Psychology in Spain, 2002, Vol. 6. No 1, 12-25

METAEVALUATION OF A TOTAL QUALITY MANAGEMENT EVALUATION SYSTEM

E. Rebolloso*, B. Fernández-Ramirez**, P. Canton and C. Pozo
University of Almería

The present paper describes the metaevaluation of a TQM evaluation system, developed for assessing the quality of management services in university administration. The evaluation follows an active and participative procedure, committed to the general values of negotiation, consensus and responsiveness. Consequently, the metaevaluation focuses on the clients' or participants' point of view in the process, in order to fit in with the original framework of the EFQM guide to the needs and perspectives of service and administration personnel. Specifically, metaevaluation criteria of comprehension, clarity, relevance, utility and effort are used. Limitations due to sample size are overcome by applying a combination of non-parametric tests in the data analysis. Results show the level of agreement reached between evaluators and participants, and suggest a certain redundance of the metaevaluation criteria and the need for review and improvement of some quality factors included in the original self-evaluation guide. Furthermore, the results highlight the relevance of context and leadership as key elements for understanding and improving administration services, as well as the scarce importance of social impact and resources management.

El presente estudio describe la metaevaluación de un sistema de Evaluación de la Gestión Integral de la Calidad, desarrollado para valorar la calidad de los servicios de administración universitaria. La evaluación sigue un procedimiento activo y participativo, comprometido con los valores de negociación, consenso y responsividad. En consecuencia, la metaevaluación se centra en la perspectiva de los clientes o participantes en el proceso, con el fin de aproximar el marco de la guía EFQM a las necesidades y perspectivas del personal de administración y servicios. En concreto, se utilizan los criterios de comprensión, claridad, relevancia, utilidad y esfuerzo. Las limitaciones muestrales se superan aplicando una combinación de pruebas no paramétricas en el análisis de los datos. Los resultados muestran los niveles de concordancia alcanzados entre evaluadores y participantes, sugieren cierta redundancia entre los criterios de metaevaluación y aconsejan la revisión y mejora de algunos factores de calidad incluidos en la guía original de autoevaluación. Se señala la importancia del contexto y del liderazgo como elementos fundamentales para comprender y mejorar el Servicio que ha sido evaluado, así como la escasa trascendencia de los factores de impacto social y gestión de recursos.

We are accustomed to hearing glowing reports of quality systems with regard to increases in productivity, reduction of errors and greater client satisfaction. Nevertheless, we are all aware that they also mean good business for some national and international organizations. There is clearly a need, therefore, for specific studies that demonstrate the reality of their effectiveness and the viability of their implementation (Choi & Ebock, 1998; de Cock & Hipkin, 1997; Westphal, Gulati & Shortell, 1997). In this work we aim to help clarify this matter by means of the concept and practice of metaevaluation, applied to the review of the evaluative process of a quality system recently introduced in a university management service (Rebolloso, Fernández-Ramírez, Pozo & Peñín, 1999; Rebolloso, Fernández-Ramírez, Pozo & Cantón, 2001). Specifically, we use the criteria of comprehension, clarity, relevance, utility and effort, in an attempt to determine the value or success of the self-evaluation guide designed, since these criteria influence the potential of the evaluation to be used directly in the improvement of services. The results of the study allow us to suggest, moreover, the importance and viability of the different factors included in the quality system (EFQM, 1999).

METAEVALUATION CRITERIA

By "evaluation", we understand scientific activity for determining the merit or value of an object (Rebolloso & Rebolloso, 1998; Stufflebeam & Shinkfield, 1987) on the basis of a series of previously-determined value cri-

---

The original Spanish version of this paper has been previously published in *Psicología del Trabajo y de las Organizaciones*, 2001, Vol. 17, No. 1, 21-43

....

Correspondence concerning this article should be addressed to E. Rebolloso. Unidad de Calidad e Investigación de Evaluación, Área de Psicología Social, Universidad de Almería, La Cañada de San Urbano, s/n, 04120 Almería. Spain. E-mail: erebollo@ual.es

teria (Scriven, 1996; Weiss, 1990). For example, we may agree with a group of clients on the importance of a service that is *rapid and error-free*; using these criteria we can evaluate the quality of different service options, and reach the conclusion that one is better than the others. Secondly, by "metaevaluation", we understand the application of this procedure to the process of evaluation itself (Chelimsky, 1983; Rebolloso, 1994a). The question is to determine which value criteria appear most appropriate, and to decide whether the process has been a suitable one from the point of view of these criteria. Finally, within this framework we carry out the metaevaluation of a Total quality evaluation system introduced in a university management service (Rebolloso, Fernández-Ramírez, Cantón & Pozo, 2000), with the aim of assessing whether the evaluation model set up is a good one according to a given set of criteria, with a view to its improvement in subsequent applications in organizational contexts.

Evaluators are professionally obliged to analyze the process they have used and the decisions they have made throughout the evaluation (Chelimsky, 1983; Rebolloso, 1994b; Scriven, 1969; Stufflebeam, 1975). The metaevaluation (or evaluation of the evaluation itself) fulfils the three most common purposes (Chelimsky, 1997; Stufflebeam & Shinkfield, 1987): a) *formative*, pursuing learning from errors, correctness of conclusions and improvement of future evaluation plans, b) *summative*, if it required to demonstrate the suitability and effectiveness of the evaluation process; and c) *knowledge-related*, if the aim is to improve the theory and practice of the discipline.

Among the theoretical frameworks available for carrying out metaevaluations, we shall describe three possibilities: the validity requirements associated with scientific practice, the standards developed by professional associations and the procedures proposed by various authors independently (M. Scriven and E. Chelimsky, in our case).

Variations on the traditional scientific perspective

Those evaluators who advocate a quantitative and experimental research model defend the "classical" scientific standards, which include the need for the research design to establish valid causal relationships (internal and external validity), through the use of valid procedures (construct validity and validity of statistical conclusion).

From standpoints that are critical of the conventional paradigm it is argued that it is more important to achieve external validity, at least in the evaluation of programs, where the utility and impact of the evaluation results are critical to the survival of the discipline (Cronbach, 1982; Cronbach & assoc., 1980; Stake, 1975). On the other hand, validity, like the notion of "objectivity", has to do with the agreement reached about what is understood by the quality of a research project, so that it becomes more a matter of intersubjectivity and consensus within the scientific community (Fernández Dols, 1990; Glass & Ellett, 1980; Kaplan, 1964; Ziman, 1981). Criticism has become accentuated with the debate on naturalist (qualitative, constructionist) evaluation procedures. Fourth-generation evaluators propose two types of criteria: *trustworthiness* and *authenticity* (Guba & Lincoln, 1989; Lincoln & Guba, 1986). The difference resides in considering research as a social construct, in which validity depends on negotiation and consensus between evaluators, the scientific community and the program's audiences (Guba, 1967; Seoane, 1996).

In the field of evaluation, where the political dimension and the necessary accommodation of the values in play (Cronbach & assoc., 1980; Weiss, 1990) have led to an awareness of the importance of the process of negotiation and search for consensus, designs should be flexible enough to adapt to the needs of those in charge of organizations and optimize the research opportunities in a changing political environment (Conner, Altman & Jackson, 1984; Rebolloso, 1990). A responsive and participatory method of work appears the most suitable for arousing the interest of the people involved in knowing and using the results of the evaluation (Patton, 1994, 1997; Rebolloso, 1994a; Rebolloso, Fernández-Ramírez, Cantón & Pozo, 2000). Moreover, no evaluator assumes the role of being in possession of any indisputable "truth" that must be accepted by others. Rather, s/he serves the interests of specific people and audiences, so that s/he must reconcile the methodological criteria of quality with the obligation to clarify and respond to their perspectives, uncertainties and needs (Rebolloso & Rebolloso, 1998)

Chen (1988, 1990) has proposed a synthesis of quality criteria for evaluation, taking into account the traditional methodological perspective and the criticisms that have emerged among evaluators of a constructivist bent. This author proposes four criteria: (a) *objectivity*, which can be likened to a criterion of reliability or confirmation (absence of subjective biases; Scriven, 1991), reached by intersubjective consensus; (b) *confidence*, which implies that the evidence provided by the evaluation is convincing or trustworthy for use by those with an interest in it; (c) *generalizability*, which asks whether the results of the evaluation can be applied to future circumstances or problems of interest to the stakeholders; and (d) *responsiveness*, which implies that the evaluation should be relevant and useful for the needs and interests of all audiences (with a view to producing actions and changes that lead to the social improvement of their members), including values such as timeliness, relevance and breadth of consequences.

Metaevaluation procedures

Professional norms include a series of prescriptions and suggestions on the requirements that should be met by evaluators in the course of their activity (Evaluation Research Society, ERS, 1981; Joint Committee on Standards for Educational Evaluation, JCS, 1981, 1988; Stufflebeam & Shinkfield, 1987). Within the wide range of existing norms, Rebolloso (1994b) recommends the use of the JCS standards, bearing in mind their wide range of views and the possibility of their application in a variety of evaluation contexts. The JCS norms are grouped in four sets: (a) *viability norms*, which demand that the evaluation be easy to implement, efficient in the use of time and resources, equipped with appropriate resources and viable from a certain number of different perspectives; (b) *integrity norms*, which require that the evaluation be in accordance with legal and ethical principles, and demand respect for privacy, freedom of information and protection of persons; (c) *accuracy norms*, which attempt to ensure the accuracy of the results and conclusions of the evaluation, avoiding any type of bias; and finally, (d) *utility norms*, aimed at guaranteeing evaluations that are informative, timely and influential, helping audiences to make better judgements or perform better than they normally did in the past.

As regards the carrying out of metaevaluations, Cronbach & assoc. (1980) recommend a review procedure in pairs, on the assumption that the honesty and balance of the evaluation will be increased through critical reviews of the performance of evaluators and sponsors. Reviews in pairs generally form part of current accreditation systems, such as that proposed by the European Foundation for Quality Management (EFQM), or those followed by university evaluation systems in several European countries (Consejo de Universidades, 1998; Van Vught et al., 1993). Nevertheless, Trow (1999; see also Cook, 1984) criticizes the “trickery” used to prepare an institution in order to present an excellent image to the visiting external evaluators, such as the last-minute hiring of teachers with brilliant CVs, exhaustive rehearsals of the visit, the use of consultants to assess how to present the institution, and so on.

Scriven (1969) argues that a variety of metaevaluation procedures can be used, including those involving the so-called checklists or professional rulebooks. Scriven’s checklist (the evaluation multimodel) comprises 21 points in the form of questions about different decisions and tasks frequently involved in the evaluation process (Stufflebeam & Shinkfield, 1987). The evaluator may go through the list several times in the course of the evaluation, analyzing whether each question has been dealt with correctly and whether it is necessary to go back in order to ensure the validity of the final result.

In a highly pragmatic approach, Chelimsky (1983) proposed a synthesis of quality criteria, distinguishing between the *technical appropriateness* and the utility of the evaluation. Technical appropriateness depends on the limitations of the work (type of information required, funds available, how soon results are required), and is defined as the difference in cost between one design and another. It can be determined whether the design is suitable taking into account the user’s needs. Technical appropriateness is relative, and demands an equilibrium between several elements: appropriateness of design, appropriateness in terms of execution, existence of viable options and absence of conceptual errors.

Utility involves four objectively analyzable components: relevance of the data with regard to informational needs, timeliness of the evaluation results, presentation of the report (understood as the logical organization of the information so that it is coherent and easily understood) and the true impact of the report (in terms of demonstrable use or its influence in legislation or decision-making). All utility measures are aimed at reinforcing the execution of specific negotiations with the client and the monitoring of satisfaction and use.

In conclusion, evaluators, especially since the so-called crisis of utilization (Cook, 1997; Patton, 1997; Shadish, Cook & Leviton, 1991), have tried to extend the traditional perspective of scientific quality, based on the concept of validity, to include a concern for the fact that evaluations achieve a clear impact in terms of the solution and improvement of social and organizational problems and conditions. Although objectivity is necessary in order to gain credibility vis-à-vis clients, it is not sufficient if the effort made is not useful for those with an interest in the evaluation process.

In spite of working from a systematic perspective (that is, attempting to bring together several issues, criteria, activities, etc., within the evaluation process; Chen, 1990; Rebolloso, 1994a; Rossi & Freeman, 1989), not all criteria can be satisfied with equal guarantees in the same study. Ideally, an agreement should be reached on which ones will be maximized, evaluating different theories in a time sequence, or balancing all of them within the program's theory, working with appropriate scientific and political values (Chen, 1990). In our case, given the novelty of evaluation processes in the context of university management, it would appear of interest to stress criteria that ensure that evaluation gains supporters and a reputation for its usefulness and advantages in the improvement of services. Consequently, we shall highlight some criteria aimed at increasing the value of evaluation for the interests of those responsible for organizations.

OBJECTIVES OF THE RESEARCH

The main objective was to review the process followed in the evaluation of a university management service (Rebolloso, Fernández-Ramírez, Cantón & Pozo, 2000; Rebolloso, Fernández-Ramírez, Pozo & Cantón, 2001; Rebolloso, Fernández-Ramírez, Pozo & Peñín, 1999). We worked with an active and participatory evaluation model that aimed, while maintaining traditional standards of objectivity and validity, to optimize decisions in order to increase responsiveness or capacity for adaptation to the informational needs of those participating in the process.

In other words, we opted for developing an evaluation tool that was directly useful for the interests of the organization, and which would serve as an aid to decisions on the improvement of services and help to achieve a positive attitude towards the work of evaluation. The fundamental reason for this choice lies in the conviction that the failure of many evaluations has resulted from their inability to deal with the real informational needs of their clients (Cronbach & assoc., 1980; Patton, 1997; Sonnichsen, 2000).

In sum, we analyze whether the process is efficient (effort), and sufficiently informative (relevance, utility), and check that there are no problems for understanding the questions posed in the self-evaluation guide (comprehension, clarity). Our aim is not to assess a theoretical structure of metaevaluation analysis, so that we have opted to select only some of the criteria potentially available. Basically, the choice is based on reasons of a pragmatic nature, bearing in mind the limitations of the participants (they are only asked questions they can answer easily and without excessive effort), and that the basic objective is to discover how to improve the guide so that is it more easily understood, requires less effort and is more relevant and useful.

The peculiar situation of our management services, with respect to the possible generalization of TQM strategies (see Rebolloso, 1999), is especially needful of metaevaluation activities. The original guides were designed to be applied in the broad context of social services and public administration (Consejo de Universidades, 1998; EFQM, 1996/97; MAP, 1999). In general, we lack theoretical guidelines to adequately comprehend the implications of the new management systems in the university administration environment, and we need to develop studies aimed at widening, trimming, tuning or modifying questions that arise in the management system and in the evaluation procedure.

In sum, the aim of the metaevaluation is, firstly, to modify the self-evaluation guide, including criteria that optimize effort and the guide's comprehension and utility, and secondly, to compare the perspective of the different participants in the evaluation process.

METHOD

Sample

The project benefited from the participation of five members of the administrative staff, constituted in a Service Quality Committee, and of two technicians from the Unit of Quality and Evaluation Research (UQER), the department responsible for evaluation activities and the setting-up of quality systems in our university. These technicians co-ordinated the self-evaluation activities. Obviously, the sample is one of convenience, by way of a case study in which all the people involved in the framework project participated. Issues of representativeness are not raised; rather, we take advantage of the opinion of all available staff. This measure does not reduce the validity of any of them as key informants who can give an opinion, from direct experience, on the strengths and weaknesses of the procedure and the self-evaluation guide.

Instruments

The Self-evaluation Guide for the university management services is made up of a set of eleven quality factors[1], derived basically from the EFQM model (1996/97, 1999). Each of them is subdivided into a set of sub-criteria or particular activities that are assumed to be associated with good organizational administration practices. The eleven factors are ordered in the following manner (Rebolloso, Fernández-Ramírez, Pozo & Cantón, 2001):

1. *Evolution and current context of the Service*. For the description and analysis of the main landmarks in the history of the Service, which have to some extent determined its current situation.
2. *Leadership*. This refers to the degree of visible commitment of those in charge of the organization and the Service in the creation and use of quality management values and strategies.
3. *Policy and strategy*. To consider how the Service defines and formulates its strategic policy, translating it into specific plans and actions that incorporate the principles and tools of quality.
4. *Personnel management*. To assess the level of participation of the Service's personnel in the achievement of improvements.
5. *Resources*. This refers to the management procedures of the main resources of the Service (financial, technological, facilities, etc.), so that they are oriented towards the objectives of quality.
6. *Processes*. How the Service manages its key processes and supporting processes, that is, how they are identified, reviewed and corrected, if necessary, in order to ensure continuous improvement.
7. *Client satisfaction*. To assess how client satisfaction is understood and increased by means of direct measures (opinion surveys) and indirect measures (complaints statistics).
8. *Personnel satisfaction*. To study how personnel satisfaction is understood and increased by means of direct measures (opinion surveys) and indirect measures (absenteeism, training levels, internal promotion, recognition of effort, rotation, etc.).
9. *Social impact*. To examine the activities of the Service that have direct and immediate consequences for the local and regional community, and for society in general.
10. *General results*. For the discussion of the results of actions under the responsibility of the Service and that are considered most significant in relation to its activity. Also, for the study of financial results (resources obtained by the Service itself, use of budget, reductions in cost of services, etc.).
11. *Metaevaluation*. Finally, a critical judgement is made of the evaluation process and suggestions for improvement are analyzed.

The instrument used for carrying out the metaevaluation was a questionnaire designed to analyze each one of the first ten factors, as well as the global work procedure. Specifically, each factor was analyzed by means of two types of item. Using an open format, we asked about the strong and weak points of each general factor, and requested suggestions for improving their questions and content. Subsequently, with a closed response format (5-point Likert scales, from “totally disagree” to “totally agree”), each of the sub-factors belonging to each general factor (41 in total) was evaluated according to five metaevaluation criteria (see below).

In sum, the metaevaluation protocol is composed of 41 sub-factors of quality (rated through closed response), which are grouped in 10 general factors (rated through open response).

Since the possible metaevaluation criteria are varied and cannot easily be satisfied at the same time (Chen, 1990), it is important to select those that best correspond to the point of view of the administration personnel, and can produce richer information, bearing in mind that the objective of adapting the self-evaluation guide to the reality of university services is more likely to be achieved by attending to the problems, doubts and needs of those involved in the process. Despite the theoretical deficiencies of the guides, there is enough previous research (Oakland, 1993; Rebolloso, 1999) and sufficiently widespread use in different types of organization to give minimum guarantees of the scientific requirements of validity and reliability, and for us to concentrate on other relevant matters.

The five criteria selected are identified with the con-

---

[1] In the current evaluation guides the term *criteria* is used to denote the broad areas of management that are considered relevant for the analysis of quality. In evaluation models *criteria* also means the variables or measures that indicate the value of the object. In order to avoid confusion, we have opted to use the term in this second sense, referring to the former concept with the term *quality factors*.

cept of utility proposed by Chelimsky (1983), and with the norms of accuracy, utility and viability of the JCS (1981, 1989), with the aim of improving the procedure of evaluation, and increasing its informative power and its capacity for influencing the system. Specifically, we started out from Chen's (1990) classification, especially in relation to the standards of objectivity and responsiveness:

- *Comprehension*. The concept involved is easily understood.
- *Clarity*. It is clearly expressed.
- *Relevance*. It is key for understanding the functioning of the Service.
- *Utility*. It is useful for improving the Service.
- *Effort*. Level of effort needed for responding to the question.

The aim of the first two (comprehension and clarity) is to obtain information in order to improve the objectivity that suggests the guide as an instrument of data-collection, in the sense that they ensure that participants understand the questions they are asked and do not perceive any kind of bias. The next two (relevance and utility) contribute to improving the responsiveness of the guide, increasing its proximity to and suitability for the needs and interests of the professional staff involved. Any of these objectives will help to promote the use of the evaluation results for improving the climate and the quality activities of the organization.

The final criterion (effort) was included as a result of the problems repeatedly indicated by participants in the evaluation. The workload of administrative staff is heavy at certain times of the year, and it should be avoided that the evaluation involves an added burden that may prove excessive, as a result of which we may expect adverse reactions or demotivation effects (Fabi & Pons, 1995; Peiró & González Romá, 1993). Effort is related, indirectly, to norms of viability (value of efficiency). In sum, the objective is that the evaluation be as easy to apply as possible (JCS, 1988).

Procedure

The initial objective of this study was to design a self-evaluation guide adapted to the particular context of university management and administration. Working from a first draft of the project, we collaborated with different management chiefs (manager, section leaders, administrative and service personnel), developing an annual project of evaluation of service quality.

The self-evaluation process required 21 work sessions over a period of approximately five months (Rebolloso, Fernández-Ramírez, Pozo & Peñín, 1999). Each quality factor was analyzed as follows. At a first meeting, the self-evaluation team discussed the questions and procedures that would be followed for satisfying the first quality factor of the guide. During the following days, the members of the Committee took responsibility for coordinating their colleagues and gathering the required information. Finally a further meeting saw the discussion of the results presented and the preparation of the final report. At all times, the two evaluators who were UQER members advocated an attitude of negotiation, seeking consensus on any matter raised, including the work procedure, the structure of the guide, the formulation of the questions and the final wording of the self-evaluation report (Rebolloso, 1994a; Rebolloso, Fernández-Ramírez, Cantón & Pozo, 2000).

The metaevaluation was considered as a task to be developed once the process had been completed, though improvement of the process was sought at all times, with the technical team adopting a responsive and negotiatory attitude vis-à-vis the participants (Guba & Lincoln, 1989; Rebolloso, 1994a; Rebolloso, Fernández-Ramírez, Cantón & Pozo, 2000). By delaying the metaevaluation we avoided overloading the committee members, since they had not been relieved in any way of their normal responsibilities within the Service.

The metaevaluation questionnaire was presented in the last work session, and its suitability, format and relevance for the improvement of the evaluation system was discussed with the participants. Its completion involved considerable effort, since it required thinking once again about each phase of the process that was about to be completed, with the added problem that some factor or area of analysis was poorly recalled, or had changed its meaning in view of the perspective acquired after obtaining an overall idea of the self-evaluation. In order to overcome participants' reluctance and the difficulties involved, the task was reduced in those cases in which it was not essential for obtaining their general impression about each quality factor. Thus, they were questioned only about the 41 sub-factors of the self-evaluation guide, and were not required to respond to each one of the specific questions of which they were composed. Nevertheless, and in spite of the negotiatory attitude of the technical team, two participants failed to return the questionnaire.

Analysis

In the first place, we measure the level of concordance between the respondents to the metaevaluation questionnaire, jointly and separately (members of the Service Quality Committee and technical team), comparing their direct responses throughout the questionnaire. We use Kendall's *W* concordance coefficient and the Mann-Whitney *U* test. The Kendall coefficient is a non-parametric test of the hypothesis that several related samples come from the same population, generally used for measuring agreement between evaluators. Kendall's *W* varied between 0 (no agreement) and 1 (total agreement). Statistical significance is calculated by means of χ^2. The Mann-Whitney test is the non-parametric equivalent of the *t* test, which checks whether two independent samples come from the same population. Statistical significance is calculated by approximation to the normal curve. Moreover, we take into account the Cronbach's alpha value, which can also be interpreted as a measure of the correlation between respondents.

Although in the text we include the results of both tests, the fact is that it is questionable to give respondents the role of judges, insofar as what interests us is not so much their degree of agreement as their opinion about the quality of each one of the sub-factors of the guide. Kendall's test measures concordance, while the alpha value measures covariance, so that we pay preferential attention to the latter in the corresponding analyses.

Subsequently, we analyze the homogeneity between the metaevaluation criteria, in order to determine whether they are independent, or whether they provide redundant information. We use the transposition of the original data matrix, so that the metaevaluation criteria can be correlated. We are aware that this procedure introduces a bias of individual differences, in the sense that we consider as independent the different responses of the same respondent who uses several times the same set of criteria to evaluate a series of questions, at the risk of increasing, therefore, the levels of covariance. This bias should be taken into account on discussing the power of this analysis. Nevertheless, we chose Spearman's *rho* coefficient, it being the most suitable for making calculations with data that do not fulfil any parametric assumption. Lastly, we use the Mann-Whitney *U* test to check the differential use of the metaevaluation criteria by the members of the Service and the technical team.

Finally, the metaevaluation data are summarized following a descriptive analysis (mean and standard deviation), in order to study the comprehension, clarity, relevance, utility and effort required in each one of the areas of management or quality factors.

RESULTS

Agreement between participants

According to Kendall's test (W = .354; χ^2 = 69.004, p =.002), the level of concordance is 35.4%, too low to consider that the seven informants maintain, in general, identical opinions in the metaevaluation. However, if we

Table 1
Concordance between participants

		Context	Leadership	Policy and strategy	Personnel management	Resources
All judges	N	7	7	7	6	6
	W	.381	.036	.242	.146	.519
	χ^2	2.667	.500	10.150	3.495	9.340
	df	1	2	6	4	3
	Sig.	.102	.779	.118	.479	.025
Members of the Service	N	5	5	5	4	4
	W	.360	.050	.188	.266	.711
	χ^2	1.800	.500	5.626	4.263	8.526
	df	1	2	6	4	3
	Sig.	.180	.779	.466	.372	.036

		Processes	Client Satisf.	Personnel Satisf	Social impact	General results
All judges	N	6	7	7	6	7
	W	.341	.640	.669	.167	.143
	χ^2	14.333	8.957	9.360	2.000	1.000
	df	7	2	2	2	1
	Sig.	.046	.011	.009	.368	.317
Members of the Service	N	4	5	5	5	5
	W	.577	.612	.744	.224	.067
	χ^2	16.157	6.118	7.444	2.235	.333
	df	7	2	2	2	1
	Sig.	.024	.047	.024	.327	.564

taken into account the score in Cronbach's alpha (α = .729), the level of covariance can be considered acceptable for analyzing each quality criterion with the average data. These values would be sufficient for continuing with the description of the data, though they raise the question of whether the differences of opinion are greater among the members of the Quality Committee and the technical team, and for which content of the guide there is greatest disagreement. The degree of agreement between the five members of the Committee (W = .406; χ^2 = 63.325, p = .008) is somewhat higher than that found for total participants, despite the fact that the Cronbach coefficient (α = .491) falls markedly. It is impossible to calculate the concordance coefficient between the two members of the technical team, even though the level of covariance between the two is greater if it is expressed by means of Cronbach's alpha (a = .670).

In order to determine the differences between groups we used the Mann-Whitney U test, comparing the ratings made for each sub-factor of the guide, and averaging the corresponding direct responses. In no case is the comparison significant, ranging from U = 5 (z = .000, p = 1.000) to U = 0 (z = -1.879, p =.060).

In order to complement these analyses, Table 1 shows the levels of concordance of the total sample of participants and of the members of the Committee, for each factor of the guide (between the two members of the technical team they could not be calculated). It can be seen that the highest levels of agreement were obtained in the sections Resources, Processes, Client Satisfaction and Personnel Satisfaction, the only ones in which there are statistically significant values.

Relationships between the five metaevaluation criteria

As already mentioned, in order to determine the empirical relationships we used the transposition of the original data matrix, from which we calculated the correlations matrix (Table 2).

The results suggest that three metaevaluation criteria could have been used, taking into account the grouping of the variables *comprehension-clarity* (r= .848, p <.001) and *relevance-utility* (r =.725, p <.001). Between the two groups there is a positive relationship, indicating that better comprehension and greater clarity of each topic evaluated in the guide is associated with ratings of greater relevance and utility for the improvement of the Service. As far as the effort variable is concerned, it is closely related to the values of comprehension-clarity. That is, the effort required is independent of whether the management areas included in the guide are more or less relevant for understanding and improving the Service, and is more related to the difficulties that arose during the evaluation process for reaching agreements on the meaning of some questions in the guide.

Metaevaluation of the factors of the Self-evaluation Guide

First of all, we examine the differential use made by the two groups of participants of the five metaevaluation criteria. To this end, we averaged their responses within each quality factor (context, leadership, policy and strategy, etc.), in order to compare the ratings by means of the Mann-Whitney U test. In general, the resulting U values indicate non-significant tendencies (these data are not included), suggesting that the two members of the technical team tend to rate the contents of the guide as more comprehensible, clearer, and more relevant and more useful for the improvement of the Service. Likewise, they state in all cases that the task requires less effort than was attributed to it by the members of the Service Quality Committee. Nevertheless, none of the results attains statistical significance, so that we shall consider the responses of all participants as similar, vis-à-vis the following analysis.

Table 3 shows the average rating for each quality factor (context, leadership, etc.), including the opinion of all the participants in the study. To support our interpretation of these results we used the Friedman test, making comparisons between the factors, in order to determine statistically which of them scored higher or lower in comprehension, clarity, relevance, etc. We shall only discuss the comparisons that were found to be significant.

Table 2
Correlation coefficients between meta-evaluation criteria[a]

	Comprehension	Clarity	Relevance	Utility
Clarity	.848 **	1.000		
Relevance	.323 **	.387 **		
Utility	.287 **	.304	.725 **	1.000
Effort	-.442	-.449	-.066	-.076

[a] Spearman's rho
** p <.001

We shall present the results in accordance with the five metaevaluation criteria:

a- *Comprehension*. Scores exceed the mean value of the scale, indicating that, in general, the questions posed by the guide are easy to understand, despite the fact that some participants expressed a less favourable opinion, as can be seen from the factors with lower average value and higher standard deviation. The questions about *social impact* are the least well understood (χ^2 =7.000 p =.008, with leadership, policy and strategy, personnel management and client satisfaction; χ^2 =6.000 p =.014, with personnel satisfaction). To a lesser extent, also in need of review is the content of the factors *processes* (χ^2 =6.000 p =.014, with leadership, and with policy and strategy) and *general results* (χ^2 =4.000 p = .046, with client satisfaction and personnel satisfaction).

b- *Clarity*. Scores exceed the mid-point of responses, indicating the general level of clarity of the guide. Comparisons between pairs show that *social impact* is the lowest-rated question in this criterion (χ^2 =6.000 p =.014, with leadership and personnel management; and χ^2 = 5.000 p =.025, with policy and strategy, client satisfaction and personnel satisfaction), together with the questions included in the analysis of *processes* (χ^2 =6.000 p =.014, with leadership, and with policy and strategy).

c- *Relevance*. In general, all the quality factors are relevant for understanding the reality of the Service, given the high ratings found. Taking into account the number of significant differences in the comparisons, the most relevant are found to be knowledge of the previous and current *context* of the Service (χ^2 =6.000 p =.014, with resources and social impact; χ^2 =5.000 p =.025, with processes; and χ^2 =3.570 p =.059, with policy and strategy) and the section *leadership* (χ^2 =7.00, p =.008, with policy and strategy; χ^2 =6.000 p =.014, with resources; and χ^2 =5.000 p =.025, with social impact). At the opposite pole, the factors *social impact* (χ^2 =6.000 p =.014, with context of the Service and with client satisfaction; χ^2 =5.000 p =.025, with leadership), and *resources management* (χ^2 =6.000 p =.014, with context of the Service, leadership and personnel management) are the least relevant.

d- *Utility*. The high scores received, in general, by all the factors of the guide indicate their potential importance for the improvement of the Service. According to the result of the comparison between pairs, none of them is considered less useful than the rest. Only the score for *leadership* stands out, with significant differences in six cases (χ^2 =6.000, p =.014, with policy and strategy, and resources; χ^2 =5.000, p =.025, with client satisfaction and personnel satisfaction; and χ^2 =4.000, p =.046, with personnel management and processes). This result provided an index of the importance attributed to the role of those in charge of the Service in the improvement of management.

e- *Effort*. All the tasks related to the analysis of each management factor involved a considerable level of effort, bearing in mind that the scores are grouped

Table 3
Meta-evaluation of the criteria of the Guide

	Comprehension		**Clarity**		**Relevance**		**Utility**		**Effort**	
	$\bar{x}$	Sx	$\bar{x}$	Sx	$\bar{x}$	Sx	$\bar{x}$	Sx	$\bar{x}$	Sx
Context	3.21	1.219	4.07	.449	4.71	.488	4.21	.636	3.07	1.304
Leadership	4.52	.572	4.33	.608	4.52	.604	4.71	.488	2.95	1.532
Policy and strategy	4.31	.514	4.08	.692	4.15	.777	4.19	.661	2.87	1.040
Personnel management	4.40	.632	4.31	.620	4.08	.630	4.31	.615	2.71	.958
Resources	4.29	.640	4.04	.781	3.29	.765	3.62	.847	2.80	1.272
Processes	3.66	.640	3.50	.627	4.14	.588	4.12	.607	3.18	.665
Client satisfaction	4.47	.766	4.28	1.026	4.38	.678	4.33	.638	2.42	.762
Personnel satisfaction	4.47	.857	4.47	.978	4.23	1.031	4.14	.813	2.47	1.069
Social impact	3.64	.735	3.73	.912	3.66	.577	4.00	.430	2.95	1.161
General results	3.71	1.216	3.92	1.090	4.00	1.000	4.50	.774	3.92	.607

around the mid-point of responses, and none approaches the value 1, which would indicate a minimum level of effort. Specifically, the analysis of the questions related to *client satisfaction* is that which appears to have needed least effort (χ^2 =5.000, p =.025, with processes; and χ^2 =3.570, p =.059, with policy and strategy). In contrast, the analysis of *general results* is the task that required most effort (χ^2 =7.000, p =.008, with client satisfaction; χ^2 =6.000, p =.014, with policy and strategy, and personnel satisfaction; χ^2 =5.000, p =.025, with social impact; and χ^2 =3.570, p =.059, with personnel management).

CONCLUSIONS

Like any other process, evaluation needs to undergo metaevaluation, with the aim of analyzing possible problems and finding solutions for its improvement (Rossi & Freeman, 1989; Scriven, 1969, 1983; Stufflebeam & Shinkfield, 1987). We failed to find convincing methods in this regard in the available self-evaluation guides, other than brief references to the advisability of reviewing the process or the suggestion to proceed by means of review by pairs (firmly rejected by Cook [1984] and by Trow [1999]). Consequently, we designed a metaevaluation questionnaire made up of a set of criteria aimed at learning more about the perspective of participants on the evaluation process that was coming to an end. The questionnaire had satisfactory results, allowing us to detect certain problems in the evaluation process and the need to reformulate some specific factors of the guide.

Agreement between participants

As the results show, there is a degree of concordance between the ratings of the seven participants in the study, though it is far from fully satisfactory. In this regard the question arises as to what may have occurred during the self-evaluation process for the work team to have failed to reach a higher level of agreement. Despite having worked together in the process, with emphasis on the fact of discussing each one of the questions to be evaluated until reaching an agreement on its meaning, it is evident that the subsequent ratings of the members of the Service Quality Committee differ from those emitted by the members of the UQER technical team. The result is understandable, assuming that the two groups start out with different conceptual maps and value perspectives on the meaning of quality management. This gives an idea of the difficulties that those responsible for university administration may encounter with regard to the comprehension and acceptance of a total quality model. In our case, the lowest levels of agreement were found in the criteria of context, leadership, policy and strategy, personnel management, social impact and general results.

Relationships between the metaevaluation criteria

The distribution of the results indicates a certain redundance between the metaevaluation criteria employed. The first group (comprehension-clarity) would designate criteria close to the value of objectivity, or absence of bias, and the second (relevance-utility), the value of responsiveness (Chen, 1990). The two groups are related, indicating that greater comprehension and clarity of the questions is linked to the perception that they are more relevant and useful for improvement of the Service. Furthermore, the level of effort is not related to variations in the relevance and utility perceived, but rather to the extent to which the contents of the guide are formulated in a clear and comprehensible way for the participants.

It is important to bear in mind that the choice of the metaevaluation criteria did not correspond to a theoretical structure that could be validated with these data, so that any inferences made are risky. The present results indicate, in a tentative way, the potential importance of the correct formulation of the evaluation questions with regard to an increase in responsiveness; that is, the negotiatory attitude should be complemented by the technical quality of the evaluation, at least as regards the perceived objectivity gained when the participants easily understand what they are asked (Chelimsky, 1983; Rebolloso, 1990).

Metaevaluation of the criteria of the Self-evaluation Guide

In general, we can conclude that the factors included in the guide were comprehensible, clear, relevant and useful, though the effort required for their analysis was greater than was desirable. Naturally, there is considerable margin for making corrections and improvements, but in no case does it appear that they should be rejected or completely modified. Below we include some observations made during the previous process of self-evaluation, with the aim of illustrating and providing a better understanding of these results. It should also be borne in mind that the comments serve to explain the specific situation of the Service evaluated, and that their capacity for generalization is reduced by the limitations of the study itself.

1- In the first place, we should analyze the problems of comprehension and clarity of three of the factors of the guide:
    a- *Social impact*. The concept of local community, used to identify a set of external beneficiaries of the Service, is particularly wide, and there is a need for studies that determine the social sectors or groups that are relevant for the university.
    b- *Processes*. The problems of comprehension may derive from the fact that the quality model represents a new way of managing the "same old things". The system involves a change in the established practices, in favour of those advocated by a Total Quality Management system (Rebolloso, 1999) –exhaustive planning by means of a participatory method of research and action.
    c- *General results*. The results of the Service were not rated, on the grounds that, given the fact that it was so recently established, the data may fail to reflect in a valid manner the efforts made during the previous year. Furthermore, the Service did not have a system for evaluating performance, and nor did the self-evaluation guide provide procedures for the setting-up of such a system.
2. Secondly, we should highlight the greater relevance and utility of the factors *context of the Service* and *leadership*, and the lower relevance of the factors *social impact* and *resources*.
    a- *Context of the Service*. The recent creation of the Service explains the importance of the previous context and the political decisions taken. Its structure and functions were defined in this period, and many of them can only be understood by taking into account the previous situation or considering the Section that previously dealt with tasks that are now linked organically to the Service. Nevertheless, and although the context serves basically to aid understanding of the Service, it is not specially useful with regard to its improvement, as can be seen in the comparison between factors.
    b- *Leadership*. Those currently in charge of the Service have played a decisive role in the definition of its present state (planning, job definitions, distribution of functions, assignment of resources, etc.), and in matters such as the level of participation and autonomy assigned to the rest of the staff. Consequently, leadership is a matter of special relevance for understanding the current situation. Moreover, the evaluation of leadership is the most useful question for the success of improvement plans, for defining and reviewing the system of quality assurance, and for encouraging the entire staff to become involved in the efforts for improvement.
    c- *Social impact*. Numerous activities are carried out for the benefit of the local community, but their true objective is to prepare the ground so that future internal users of the Service (future students, relatives, contractors, etc.) are made more aware of what it offers and how to take advantage of it. However, there is no awareness of the Service's potential for carrying out activities that support local development. The community is considered not so much a recipient of services than as an intermediary with future users. This may be the reason why social impact is also considered relatively unimportant for improvement of the Service.
    d- *Resources*. University management services in Spain have only limited involvement in resources management, which is mainly the responsibility of other departments, among them the Rectorate. Their power in this regard is restricted to demanding resources of various types according to the internal needs detected; they do not even have contact with suppliers, who work directly with the higher authorities. For this reason, resources management, as mentioned in the guide, does not form part of the reality of the Service, and is irrelevant to an understanding of its current situation and internal management practices. It is another matter whether or not it appears useful for improvement of the Service, and we have seen that it is not considered any less so compared to other factors.
2. As regards the effort required, we already saw in the self-evaluation, on numerous occasions, the problems of participants to combine their internal res-

ponsibilities with the evaluation task. Level of effort required should be reduced in all the factors of the guide (modifying the procedure or requirements of the work). Specifically, the criterion *client satisfaction* was that which required least effort, whilst that which required most was *general results*.

a- *Client satisfaction*. The UQER carried out the study of client satisfaction and professional personnel satisfaction (Rebolloso, Fernández-Ramírez, Pozo et al., 2000), and the self-evaluation team devoted little time to the analysis of this matter, since it was sufficient to provide a minimal description of the few previous experiences and a general analysis of the results of the study of satisfaction. Evidently, in other cases the effort required will depend on the procedure and instrumentation developed for obtaining information on clients.

b- *General results*. Two problems should be taken into account. First, that several sessions were employed at the beginning of the self-evaluation process to gather the information as suggested in the Universities Council guide (Consejo de Universidades, 1998); as this information was not eventually used, the perceived effort may have been greater. And second, that the system of minimum indicators subsequently developed required the rapid collection of information that was more or less accessible in many cases, but that needed a certain degree of elaboration and organization, thus increasing the perception of effort employed.

LIMITATIONS AND FUTURE PERSPECTIVES

First of all, we should recognize the limitations of the study, taking into account above all the small sample size and the difficulties involved in the data analysis. The ideal situation is to carry out the metaevaluation once the quality system has been extended to all units of the management structure.

This study was a first rehearsal for testing the evaluation instruments and procedure, with no aspiration other than to improve them in a gradual and considered way. Thus, the conclusions are not intended to be generalizable. On the contrary, we recognize that their main utility concerns the improvement of the specific process of establishing a quality system in the context of our own university, while not discarding the possibility that the study may encourage other researchers to carry out metaevaluations as activities of critical self-analysis aimed at improving their own work.

Even so, during the metaevaluation process there appeared two important problems that demonstrate the need to improve the procedure. In the first place, the number of metaevaluation criteria could be reduced in future applications, given the conceptual and empirical proximity of some of them (comprehension-clarity; relevance-utility). Secondly, throughout the self-evaluation process, the metaevaluation should be complemented by, for example, a questionnaire or registration form to be filled out at the end of each evaluation session, or after the analysis of each quality factor. This suggestion derives from one of the problems that arose: the effort involved in having to review all the self-evaluation elements once more, just when the (longed-for) end of the process was in sight after five months of unrewarded work (we can only assume that this was why some failed to return the questionnaire).

Regardless of the political realities and the interests created within the university organization, we feel obliged to defend an ideal committed to the improvement and reinforcement of persons and institutions (Blanco, 1998; Miller, 1969; Rebolloso, 1999; Stake, 1975). Conscious of the importance of the changes demanded by the setting-up of an extensive system for the assurance of quality, and that any type of change is to some extent traumatic, we can only advise prudence, reflection and political skill for reaching the necessary consensus between the interest groups of the institution, and beginning a gradual process of change (Rebolloso, 1999).

This first self-evaluation trial has served to calibrate the evaluation procedure and instruments, and to make an initial approach to the management systems employed by university administration services. Nevertheless, the results provide an inadequate picture for understanding the reality of the whole system and for making suggestions that are generalizable in the current situation. In consequence, it is necessary to continue from this initial approach, extending the evaluation to other services in order to increase the amount of information available on the context of the introduction of quality systems.

It is advisable, finally, for subsequent evaluations to be preceded by a period for discussing and negotiating with the participating services, so that the effort required is reduced and a more valid approach to the reality of the organization is achieved. In any case, there is clearly a need to carry out systematic metaevaluations, which may be based on a questionnaire such as that used in this study, or may be broadened in accordance with the recommendations of various authors (Chelimsky, 1983; Chen, 1988, 1990; JCS, 1981, 1988; Rebolloso, 1994b; Scriven, 1969).

REFERENCES

Blanco, A. (1998). La coherencia en los compromisos. In I. Martín-Baró, *Psicología de la liberación*. Madrid: Trotta.

Chelimsky, E. (1983). The definition and measurement of evaluation quality as a management tool. *New Directions for Program Evaluation*, *18*, 113-126.

Chelimsky, E. (1997). The coming transformations in Evaluation. In E. Chelimsky & W.R. Shadish (Eds.). *Evaluation for the 21st century. A handbook* (pp. 1-26). Thousand Oaks, CA: Sage.

Chen, H.T. (1988). Validity in evaluation research: A critical assessment of current issues. *Policy and Politics*, *16*, 1-16.

Chen, H.T. (1990). *Theory-driven evaluations*. Newbury Park, CA: Sage.

Choi, T.Y. & Eboch, K. (1998). The TQM paradox: Relations among TQM practices, plant performance, and customer satisfaction. *Journal of Operations Management*, *17*, 59-75.

Conner, R.F., Altman, D.G. & Jackson, C. (1984). 1984: A brave new world for evaluation. In R.F. Conner, D.G. Altman & C. Jackson (Eds.). *Evaluation Studies Review Annual* (Vol. 9). (pp. 12-22). Beverly Hills, CA: Sage.

Consejo de Universidades (1998). *Plan nacional de evaluación de la calidad de las universidades*. Madrid: Secretaría General del Consejo de Universidades.

Cook, T.D. (1984). Opportunities for evaluation in the next few years. In R.F. Conner, D.G. Altman & C. Jackson (Eds.). *Evaluation Studies Review Annual* (Vol. 9). (pp. 726-751). Beverly Hills, CA: Sage.

Cook, T.D. (1997). Lessons learned in evaluation over the past 25 years. In E. Chelimsky & W.R. Shadish (Eds.). *Evaluation for the 21st century*. *A handbook* (pp. 30-52). Thousand Oaks, CA: Sage.

Cronbach, L.J. (1982). In praised of uncertainty. In P.H. Rossi (Ed.). *Standards for evaluation practice* (pp. 49-58). San Francisco, CA: Sage.

Cronbach, L.J. & assoc. (1980). *Toward reform of program evaluation*. San Francisco, CA: Jossey-Bass.

De Cock, C. & Hipkin, I. (1997). TQM and BPR: Beyond the beyond myth. *Journal of Management Studies*, *34*, 659-675.

Deming, W.E. (1986). Japan: Quality control and innovation. *Business Week*, *2697*, 18-44.

European Foundation for Quality Management (1996/1997). *Premio europeo a la calidad*. Madrid: Club de Gestión de la Calidad.

European Foundation for Quality Management (1999). *Modelo EFQM de excelencia*. Madrid: Club de Gestión de la Calidad.

Evaluation Research Society (1981). *Standards for program evaluation*. Potomac, MD: Evaluation Research Society.

Fabi, B. & Pons, O. (1995). Los círculos de calidad: la experiencia internacional. *Alta Dirección*, *182*, 275-282.

Fernández Dols, J.M. (1990). *Patrones para el diseño de la Psicología social*. Madrid: Morata.

Glass, G.V. & Ellet, F.S. (1980). Evaluation research. *Annual Review of Psychology*, *31*, 211-228.

Guba, E.G. (1967). Evaluation and the process of change. In *Notes and working papers concerning the administration of programs: Title III ESEA*. Washington DC: U.S. Senate Committee on Labor and Public Welfare.

Guba, E. G. & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.

Joint Committee on Standards for Educational Evaluation (1981). *Standards for evaluations of educational programs*, *projects and materials*. New York: McGraw-Hill.

Joint Committee on Standards for Educational Evaluation (1988). *The personnel evaluation standards: How to assess systems for evaluating educators*. Newbury Park, CA: Sage.

Kaplan, A. (1964). *The conduct of inquiry*. San Francisco, CA: Chandler.

Lincoln, Y.S. & Guba, E.G. (1986). But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. In D.D. Williams (Ed.). *Naturalistic Evaluation. New Directions for Program Evaluation* (Vol. 30). San Francisco, CA: Jossey-Bass.

Ministerio de Administraciones Públicas (1999). *Guía de autoevaluación para la administración pública*. Madrid: MAP/BOE.

Miller, G.A. (1969). Psychology as a mean of promoting human welfare. *American Psychologist*, *24*, 1063-1075.

Oakland, J.S. (1993). *Total quality management. The route to improving performance*. Oxford: Butterworth-Heinemann.

Patton, M.Q. (1994). Developmental evaluation. *Evaluation Practice*, *15,* 311-320.

Patton, M.Q. (1997). *Utilization-focused evaluation*. Newbury Park, CA: Sage.

Peiró, J.M. & González Romá, V. (1993). *Círculos de calidad*. Madrid: Eudema.

Rebolloso, E. (1990). Algunos apuntes sobre la utilidad de la evaluación en el marco de la toma de decisiones políticas. In *Actas del III Congreso Nacional de Psicología Social* (pp. 567-574). La Coruña: Tórculo.

Rebolloso, E. (1994a). La evaluación del personal profesional: ¿Un modelo de evaluación nuevo o una síntesis de modelos previos? *Revista de Psicología Social*, *2,* 231-255.

Rebolloso, E. (1994b). Los standards de evaluación de la Joint Committee. Su utilidad como guía para la práctica profesional en educación y servicios sociales. *Revista de Psicología Social Aplicada*, *4*, 63-91.

Rebolloso, E. (1999). *La evaluación de la calidad como estrategia de supervivencia y futuro de la universidad*. Almería: Universidad de Almería.

Rebolloso, E., Fernández-Ramírez, B., Cantón, P. & Pozo, C. (2000). El papel de la investigación cualitativa en la evaluación de los servicios universitarios. *Cuadernos IRC*, *4,* 65-82.

Rebolloso, E., Fernández-Ramírez, B., Pozo, C. & Cantón, P. (2001). *Estrategias de calidad en la Universidad. Guía para la autoevaluación de los servicios de administración universitarios*. Valencia: Promolibro.

Rebolloso, E., Fernández-Ramírez, B., Pozo, C. & Peñín, R. (1999). *Autoevaluación del Servicio de Alumnos - 1999. Informe de resultados*. Almería: UCIE.

Rebolloso, E., Fernández-Ramírez, B., Pozo, C. et al. (2000). *Gestión Integral de la Calidad. Satisfacción de los usuarios y del personal profesional con los servicios universitarios*. Almería: UCIE.

Rebolloso, E. & Rebolloso, J.R. (1998). Significado y desarrollo actual de la evaluación de programas. In E. Rebolloso (Ed.). *Evaluación de programas. Ámbitos de aplicación* (pp. 219-240). Barcelona: Textos Universitarios Sant Jordi.

Rossi, P.H. & Freeman, H.E. (1989). *Evaluación. Un enfoque sistemático para programas sociales*. México: Trillas.

Scriven, M. (1969). An introduction to metaevaluation. *Educational Product Report*, *2*, 36-38.

Scriven, M. (1983). Evaluation ideologies. In G.F. Madaus, M. Scriven & D.L. Stufflebeam (Eds.). *Evaluation models* (pp. 229-260). Boston, MA: Kluwer-Nijhoff.

Scriven, M. (1991). Beyond formative and summative evaluation. In G.W. McLaughlin & D.C. Philips (Eds.). *Evaluation and education: At quarter century*. Chicago, IL: University of Chicago Press.

Scriven, M. (1996). Types of evaluation and types of evaluators. *Evaluation Practice*, *17*, 151-161.

Seoane, J. (1996). El escenario postmoderno de la psicología social. In G. Collier, H.L. Minton & G. Reynolds (Authors). *Escenarios y tendencias de la psicología social* (pp. 13-30). Madrid: Tecnos.

Shadish, W.R., Cook, T.D. & Leviton, L.C. (1991). *Foundations of program evaluation*. Newbury Park, CA: Sage.

Sonnichsen, R. C. (2000). High impact internal evaluation: A practitioner’s guide to evaluating and consulting inside organizations. Thousand Oaks, CA: Sage.

Stake, R.E. (1975). *Evaluating the arts in education: A responsive approach*. Columbus, OH: Merril.

Stufflebeam, D.L. (1975). *Metaevaluation*. Kalamazoo, MI: Evaluation Center, Western Michigan University.

Stufflebeam, D.L. & Shinkfield, A.J. (1987). *Evaluación sistemática*. Barcelona: Paidós/MEC.

Trow, M. (1999). Confianza, mercados y rendición de cuentas en la educación superior. *Cuadernos IRC*, *2*, 47-63.

Van Vught et al. (1993). *Gestión de la calidad y garantía de calidad en la enseñanza superior europea: Métodos y mecanismos*. Brussels: Publications Office of the European Communities.

Weiss, C.H. (1990). *Investigación evaluativa*. México: Trillas.

Westphal, J.D, Gulati, R. & Shortell, S.M. (1997). Customization or conformity? An institutional and network perspective on the content and consequences of TQM adoption. *Administrative Science Quarterly*, *42*, 366-394.

Ziman, J. (1981). *La credibilidad de la ciencia*. Madrid: Alianza.

Psychology in Spain, 2002, Vol. 6. No 1, 26-40

COPING STRATEGIES IN PSYCHOTICS: CONCEPTUALIZATION AND RESEARCH RESULTS

Salvador Perona Garcelán and Antonio Galán Rodríguez
Health Service of Andalusia

En el presente trabajo se revisan las investigaciones llevadas a cabo para estudiar las conductas espontáneas de afrontamiento en pacientes psicóticos. Se hace un análisis del uso que se ha hecho en dichos trabajos del concepto de afrontamiento y se describen las principales categorías de afrontamiento utilizadas en la literatura, así como de los resultados más relevantes de estos estudios. Asimismo se concluye con aportaciones prácticas para el desarrollo y planificación de estrategias de intervención en este ámbito.

This paper reviews research carried out on spontaneous coping behaviour in psychotic patients. It describes the use of the concept of coping in these studies and the major categories of coping used in the literature and presents relevant results. Finally, we include some practical contributions for the development and planning of intervention in this field.

The active nature of the individual in the face of difficulties resulting from illness is a reality that has been confirmed at both a popular and scientific level. Scientists have looked into the phenomenon with the aim of studying the different behaviours used by patients to deal with their disorders and the limitations deriving from them. But the great difficulty we encounter in this field is the multitude of approaches made, so that it is difficult to arrive at a global and comprehensive view of research on how the individual acts in these situations.

The behaviours in question are often labelled as “coping” behaviours, a concept with a long history that is currently associated with the work of Lazarus & Folkman (1986) within the framework of their Transactional Theory of Stress. These authors defined a concept that at an intuitive level was highly accessible, and which they endowed with a strict theoretical formulation that facilitated its use at a scientific level. However, despite the heuristic value of this model and its consequent popularization, the concept of coping continues to be quite loose, so that the same label is used to refer to quite diverse processes, ranging from patterns of neuroendocrinal activity to specific types of cognitive processing (Crespo & Cruzado, 1997; López-Roig, 1991).

In our own field of work, that of psychosis, there is clearly a reproduction of the situation that we describe elsewhere in relation to the concept of illness in general (Galán Rodríguez, 2000). Thus, in the studies published to date on coping in people diagnosed with psychosis, it is found that a high percentage of subjects *report* not remaining passive in the face of everyday difficulties and problems, and that they put into practice a series of strategies with the aim of feeling better. More specifically, it was found that the percentage of subjects claiming to use coping strategies ranged from 72% (e.g., Tarrier, 1978; Dittmann & Schüttler, 1990) to 100% (e.g., Böker, Brenner, Gerstner, Keller, Müller & Spichtig, 1984; Brazo, Dollfus & Petit, 1995), with the exception of the work of Carr (1988) and of Carter, Mackinnon & Copolov (1996), in which lower figures appear: 50.2% and 68% of patients, respectively.

If we review research on coping skills in individuals diagnosed as psychotic, we can see that this concept has been given a variety of names since the publication of the first empirical work by Falloon & Talbot (1981) and Lange (1981). Thus, after Falloon & Talbot’s (1981) *coping strategies* and Lange’s (1981) *coping reactions*, we find *self-control behaviour* (Breier & Strauss, 1983), *self-healing strategies* (Böker, Brenner, Gerstner, Keller, Müller & Spichtig, 1984), *autoprotective efforts* (Brenner, Böker, Müller, Spichtig & Würgler, 1987), *self-help techniques for auditory hallucinations* (Frederick & Cotanch, 1994) and *anti-hallucinatory strategies* (Brazo, Dollfus & Petit, 1995).

We are faced, then, with a wealth of concepts restricted to specific areas of functioning (such as hallucinations) or to particular behaviours (such as those of self-control). The objective of this work is to undertake a review of these varied partial approaches, in an attempt to achieve an overview of that which can be grouped under the label “coping in psychosis”.

The original Spanish version of this paper has been previously published in *Clínica y Salud*, 2002, Vol. 12 No 2, 145-178

....

Correspondence concerning this article should be addressed to Salvador Perona Garcelán. *Unidad de Rehabilitación de Area Virgen del Rocío, Avda. Kansas City, 32E bajo, 41007 Sevilla. Spain.*
E-mail: sperona@correo.cop.es

THE CONCEPT OF COPING IN THE FIELD OF PSYCHOSIS

Conceptions of coping in the study of psychoses

Despite the diversity of denominations, such as those listed above, in this work we shall employ that which is most commonly accepted and most frequently used, which is that of coping. Nevertheless, they all refer to the personal resources used by people diagnosed as psychotic to deal with the demands resulting from their disorder and those arising from their environment. Even so, we have found some differences in the conceptualization of coping among the different authors that have studied the matter.

There is one group, perhaps the most numerous, that approaches the study of coping from an empirical point of view, and on the basis of the vulnerability-stress model of Zubin & Spring (1977) and Nuechterlein & Dawson (1984). This model defends the hypothesis that all individuals with a psychotic disorder are predisposed, or vulnerable to experiencing a crisis when they come into contact with stressful situations that overwhelm them and cause the psychotic symptoms to emerge. However, this model also states that the vulnerable individual is not helpless in the face of aggression from the environment, postulating the existence of a series of protective variables that can avoid the subject having a relapse. Among these variables are biological protectors, such as anti-psychotic drugs, social ones, such as social support networks, and personal ones, such as general skills used by the subject to adapt to his/her medium (social skills, labour skills, assertive skills, instrumental skills, etc.).

Over the last twenty years, this model has stimulated a great deal of research, leading to the development of strategies for helping such people to improve their quality of life. Among these strategies are all of those described as “coping strategies”, such as training in social skills, communication skills for the families of schizophrenics, and so on. However, these strategies have emerged from the experience and research of clinical psychologists and psychiatrists that work in this field, and might therefore be considered “artificial”, in the sense that they are designed by mental health professionals and are in many cases strange for subjects, especially when they fall outside of their ideological or cultural framework, or, more specifically, are at odds with the basic set of beliefs that allow an individual to make sense of his or her reality. For this reason, many researchers, on conceptualizing psychoses from the vulnerability-stress model, have considered the importance of studying “natural” or “spontaneous” coping skills.

This was the framework of the first studies on coping in relation to psychotic symptoms (Falloon & Talbot, 1981; Tarrier, 1987; Carr, 1988), a framework that has had a great deal of influence in the field. Nevertheless, except in the case of Yusupoff & Tarrier (1996), we have found no clear definition of what researchers understand by coping. These authors define it as “the active self-generation of cognitive and behavioural procedures for directly influencing symptoms or reducing the resulting anxiety” (Yusupoff & Tarrier, 1996, p. 86). However, according to this use of the term, coping refers to all the cognitive and behavioural resources patients employ to defend themselves against the symptoms and against other intrusive experiences.

In this regard, it is highly important to take into account several elements that are implicit in these studies. In first place, we should stress the importance of detecting the antecedents of symptoms, which allows the subject to employ these behaviours “consciously” (Falloon & Talbot, 1981; Breier & Strauss, 1983; Tarrier, 1987; Carr, 1988). Secondly, and in contrast to Lazarus & Folkman’s (1986) Transactional Theory of Stress, the concept of coping is closely linked to the effectiveness of the strategies. These should be useful and should achieve, for example, the objective of reducing or eliminating the interference of the voices (Falloon & Talbot, 1981). Thirdly, of priority importance is a conceptualization of coping from a defensive perspective, that is, the subject performs such behaviours to defend him/herself from the symptoms, understanding these phenomena from a medical perspective, as intrinsically negative and the expression of an underlying psychiatric pathology. Lastly, another basic idea that emerges from these studies is the consideration of coping as a molecular behaviour, contingent upon the problem behaviour and limited in time, and whose objective is the temporary elimination of the symptom.

There are two other approaches in the literature on the field which, from a different perspective, conceptualize coping strategies as complex processes, not limited in time as in the approach described above.

The first of these is that of Romme & Escher (1989, 1996) and Romme, Honig, Noorthoorn & Escher (1992). These authors developed their research with the specific aim of studying coping strategies in situations of auditory hallucinations. Their approach is quite different from that of the authors mentioned so far, since they understand coping as a process that facilitates the integration of the hallucinatory experience in the patient’s everyday life. On not considering the voices as a pathognomonic symptom of schizophrenia or as something necessarily negative, they argue that coping does not have to be defensive, but that it can rather be associated with the search for some kind of peaceful accommodation and acceptance of the voices as “part of oneself”.

In order to understand this approach it is highly important to consider the patient’s beliefs or frames of reference in relation to the voices. Romme & Escher (1989) have found that the coping process is complex and may vary according to these frames of reference. Their function consists in providing the voices with a meaning within the individual’s life, thus making them a potentially decisive factor in the degree of adaptation to the voices.

The other approach is found in a group of authors that study coping strategies in psychotic patients, employing concepts from theories developed especially for the study of coping in other disorders and populations. Specifically, we are referring to the application of the psychotic disorder to Lazarus & Folkman’s theory (1986) (e.g., Böker, Brenner, Gerstner, Keller, Müller & Spichtig, 1984; Brenner, Böker, Müller, Spichtig & Würgler, 1987; Thurm & Haefner, 1987; Wiedl & Schötter, 1991; Wiedl, 1992).

From this perspective coping is defined as “those constantly changing cognitive and behavioural efforts developed for managing specific external and/or internal demands that are evaluated as excessive or overwhelming for the individual’s resources” (Lazarus & Folkman, 1986, p. 164). These authors formulated a transactional model of coping, which stresses the bidirectional relationship between the person and his/her environment. The main assumption of this model is that an event is not in itself stressful, but that its importance is determined by the meaning the individual attributes to it on the basis of cognitive judgement processes. Two types of judgement are distinguished: the primary judgement relates to the subject’s evaluation of the event, and the secondary one relates to the subject’s evaluation of his or her resources for coping with the event, in the case of its being considered threatening or dangerous.

A characteristic of this conception is that coping is considered as a process closely related to the contexts in which the problem appears. From this perspective, coping is not a trait, but rather a constantly changing state that evolves according to the current demands of the individual’s relationship with his/her environment or him/herself, especially when the objective of the efforts is the psychotic symptoms themselves or the basic cognitive disorders caused by the illness.

Finally, we should also highlight the fact that Lazarus & Folkman’s definition of coping takes into account the effort necessary to manage stressful demands, regardless of the result. The quality of a strategy (its effectiveness or suitability) is determined solely by its effects in a given situation and in the long term (Lazarus & Folkman, 1986).

Taxonomies of coping strategies

Attempts to classify the various behaviours an individual would use in coping processes have resulted in some interesting categorizations. Before presenting them, it is appropriate to make some observations about the way the authors have developed these taxonomies.

Thus, first of all we should draw attention to some terminological aspects of these classifications. In this regard it is important to underline the distinction that can be established between, on the one hand, *coping behaviours* and, on the other, *coping strategies*. The former term is employed for referring to those molecular, observable and quantifiable behaviours that subjects employ, consciously or otherwise, in order to protect themselves from the demands of the environment or of their own illness; that of coping strategies is used to refer to those behaviours subjects employ (consciously or not) in a planned and organized way to protect themselves from the demands of the environment or of their own illness. However, this distinction is quite difficult to make in practice, since the term strategy assumes a degree of abstraction, introspection and verbal fluency that many patients diagnosed as psychotic lack when they are required to report to researchers. Therefore, this second concept is an abstraction of the researcher him/herself, who simply categorizes the behaviours described using a verbal term of a higher order. For example, if a patient says that on hearing voices he sometimes goes to sleep and other times relaxes, the researcher concludes that the patient utilizes coping strategies aimed at the reduction of arousal.

Secondly, we should mention the methodological aspects present in the development of these taxonomies. In the review we have made, nearly all the studies use semi-structured interviews, organized in two phases: the first is normally designed to explore the problems presented by the patient (psychotic symptoms, prodromes, stressful life events, etc.), and on the basis of the information obtained there are two alternatives: a) to formulate open questions in which the patient is asked directly what s/he does when faced with each of the problems described in the first phase; the researcher records the frequency of the specific coping behaviours and classifies them according to general coping strategies groups (e.g., Falloon & Talbot, 1981; Cohen & Berk, 1985 Tarrier, 1987); and b) to provide the patient with a predetermined list of coping behaviours, generally based on previous studies, for them to identify those they normally use (e.g., Carr, 1988; O’Sullivan, 1994; Carter et al. 1996); a few studies have used standardized questionnaires developed from the study of coping in other populations (e.g., Farhall & Gehrke, 1997; Van Den

Bosch & Rombouts, 1997; MacDonald, Pica, McDonald, Hayes & Baglioni, 1998).

The taxonomies of coping described in the literature are quite varied, and there is no consensus between authors with respect to the matter. The most suitable way of describing them, with a view to their understanding, is to classify them in two groups: structural and functional.

The commonest approaches are those of a structural nature. They consist in descriptions of what the subject is doing at a given moment vis-à-vis certain situations judged as problematic. These descriptions are made on the basis of behaviour, and are grouped according to arbitrary and pragmatic criteria, without recourse to a specific theory on coping. Although these taxonomies are quite varied, the most frequently used are those made according to the classical topographical analysis traditionally employed in the cognitive-behavioural evaluation of any behavioural problem. In these cases coping strategies are categorized as behavioural, physiological and cognitive (Falloon & Talbot, 1981; Frederick & Contanch, 1994; Brazo, Dollfus & Petit, 1995). A number of authors include, together with the above categories, some others, such as sensorial strategies (Tarrier, 1987), social coping (Carr, 1988, Mueser, Valentine & Agresta, 1997), or coping based on medical strategies and on the symptoms themselves (Carr, 1988). Other studies use more general classifications, obtained on the basis of observation and clinical practice, and which could be included in any of the groupings previously described. For example, Breier & Strauss (1983) made a classification in three groups: self-instructions, reduction of activity and increase of activity, while McNally & Goldberg's (1997) classification covers only the context of cognitive strategies (logic and reason, objectivization, substitution, distraction, etc.) (see Figure 1).

A problem presented by these taxonomies is the arbitrary nature of the classification criteria, such that we may find the same coping tactic or behaviour included in different strategies. A clear example is provided by distraction: for Falloon & Talbot (1981) it is a cognitive strategy, but for Carr (1988) it is a behavioural strategy.

Within the strategies we have labelled structural, an important body of research has been contributed by a group of authors that have applied the mathematical technique of factorial analysis (O'Sullivan, 1994; Carter et al. 1996; Farhall & Gehrke, 1997). Specifically, they have done so in relation to auditory hallucinations, where there is also great disparity between different classifications (see Figure 2).

Carter et al. (1996) identified, in the factorial analysis of the 26 coping strategies used by the subjects in their study, three factors that explained 81% of the variance of their data. As Figure 2 shows, the three factors ("strategies based on subvocal speech", "search for competing auditory stimulation", and "well-integrated or intellectual responses to an intrusive stimulus") are similar to other groupings of an arbitrary nature made by authors already mentioned. For example, these three factors remind us of Tarrier's (1987) taxonomy, and at the same time coincide with the results of the Slade & Bentall's (1989) research on the key factors (distraction, reduction of anxiety and focalization) that describe the positive results of cognitive-behavioural interventions with auditory hallucinations (for an extensive review of this matter, see Perona Garcelán & Cuevas Yust, 1996).

However, the results of the factorial analysis by O'Sullivan (1994) and Farhall & Gehrke (1997) are quite different from those of the above study. The factors that include the different coping strategies in these studies have in common the fact of being based on the final result of such strategies (e.g., "hopeful and optimistic engagement", "despairing rejection", "ambivalent acceptance" and "hopeful rejection" – O'Sullivan, 1994).

The second group of strategies is that which we have called functional. Studies that have used strategies of a functional type are those coming from the research tradition of Lazarus & Folkman (1986). An important characteristic of this approach is that coping is conceived according to its role in the process of adaptation to the environment, avoiding the confusion between objectives and results. While the function of coping is related to the objective of each strategy, the result refers to the effect of the strategy.

From this perspective there are two basic functions of coping that allow us to make a classification of strategies in two groups: a) problem-oriented coping, which refers to behaviours oriented to manipulating or altering the problem; and b) emotion-oriented coping, which includes those behaviours whose function is that of regulating the emotional response elicited by the problem (Folkman & Lazarus, 1980).

In the study of subjects diagnosed as psychotic, this taxonomy has also been employed in several studies in relation to coping in basic cognitive disorders (Böker, Brenner et al. 1984; Brenner, Böker et al. 1987; Böker, Brenner et al. 1989; Takai et al. 1990; Wiedl & Schötter, 1991; Wiedl, 1992), as well as to positive and negative psychotic symptoms (Mueser, Valentine & Agresta, 1997; Middelboe & Mortensen, 1997; MacDonald et al. 1998).

In this research context, the terminology has varied slightly, with the use of the expressions "compensatory efforts oriented to problem-solving" and "compensatory efforts not oriented to problem-solving" to refer to the categories proposed by Folkman & Lazarus (1980) – *problem-oriented coping* and *emotion-oriented coping*, respectively. The former concept has been defined by Brenner, Böker et al.(1987) as conscious autoprotective efforts directly oriented to confronting the source of the disorder; the latter as conscious efforts to deny, isolate oneself or distance oneself from – in general, to avoid – the emotional consequences of the disorder (see Figure 3).

The level of conceptual development in these studies has been clearly higher, due perhaps to the legacy of work in the study of coping on other areas, such as that of physical health, and in other populations within the field of mental health. An example is provided by the combination of structural and functional criteria in authors such as Wiedl & Schötter (1991), Wiedl (1992), Mueser, Valentine & Agresta (1997) & Middelboe & Mortensen (1997), in whose work two axes are established for the classification of coping strategies. The first of these, which they call *coping level*, includes the structural categories of Carr (1988) or Falloon & Talbot (1981); the second axis, which they refer to as *coping direction*, includes the functional taxonomies of Folkman & Lazarus (1980 – see Figure 3). Thus, in these studies we have a taxonomy of coping strategies made up of a double-entry matrix that allows us to study them from a multidimensional perspective, and to contribute date that are highly relevant to the field.

AREAS OF INTEREST

Coping strategies most commonly used by people diagnosed as psychotic

Since the earliest work, the major concern of the different authors was to demonstrate that patients diagnosed as psychotic were capable of coping with their own pro-

Figure 1
Structural coping strategies

FALLOON & TALBOT, 1981 **FREDERICK & CONTANCH, 1994** **BRAZO, DOLLFUS & PETIT, 1995**	**BREIER & STRAUSS, 1983**	**COHEN & BERK, 1985**	**TARRIER, 1987**
Change in behaviour Change in physiological arousal Cognitive methods	Self-instructions Reduction of activity Increase of activity	Fight back Time out Individual diversion Social diversion Praying Medical strategies Drugs/alcohol Helplessness Acceptance	Cognitive strategies Behavioural strategies Sensorial strategies Physiological strategies
THURM & HAEFNER, 1987	**CARR, 1988**	**ROMME & ESCHER, 1989**	**KUMAR, THARA & RAJKUMAR, 1989**
Asking for help Intrapsychic coping Taking extra medication Behavioural change	Behavioural control Cognitive control Socialization Medical control Symptomatic behaviour	Distraction Ignoring the voices Selective listening Putting limits on the voices	Internal dialogue Speaking to a close relative or friend Seeking psychiatric help Adjusting the medication Occupational activity
DITTMANN & SCHÜTTLER, 1990	**CHADWICK BIRCHWOOD, 1994**	**McNALLY GOLDBERG, 1997**	**WAHASS & KENT, 1997**
Distancing Increase in interpersonal contact Cognitive control Symptomatic behaviour Adjustment of medication Helplessness	Resistance to the voices Commitment to the voices Indifference to the voices	Logic and reason Objectivity Replacement Distraction Change of attitude Confirmation of reality Volition Self-affirmation Humour	Religious strategies Distraction Physiological strategies Social change Individualist strategies Strategies to stop voices

blems, but they were also keen to identify which strategies were most frequently used. In the previous section we showed that these individuals report not being passive in the face of their difficulties, and we described the taxonomies found in the different studies. As the reader will recall, we classified these coping strategies as structural or functional. In accordance with this classification, we shall continue by presenting the results of research on this topic.

From a structural point of view, we have found that strategies of a behavioural type are the most frequently used (e.g., Breier & Strauss, 1983; Kanas & Barr, 1984; Carr, 1988; Takai, Uematsu, Kaiya, Inoue & Ueki, 1990; Wiedl & Schötter, 1991; Kinoshita, Yagi, Inomata & Kanba, 1991; Yagi, Kinoshita & Kanba, 1992; Middelboe & Mortensen, 1997; Pallanti, Quercioli & Pazzagli, 1997), examples being the carrying out or reduction of occupational-type activities, watching television, going for walks, doing sport, playing a musical instrument, etc. Other authors report that behavioural strategies are the most commonly used, along with cognitive or social strategies (Tarrier, 1987; Kumar, Thara & Rajkumar, 1989; Mueser, Valentine & Agresta, 1997), such as the use of self-instructions, listening attentively to the voices, stop thinking, talking to a friend, going out with someone, etc. Finally, there is a small group of studies in which it is difficult to decide which type of strategy is most frequently used, given the great variety in subjects' responses to the different stressors (Falloon & Talbot 1981; O'Sullivan, 1994; Wahass & Kent, 1997). For example, Falloon & Talbot (1981) found that the subjects in their study used, as strategies for coping with the voices, mainly the following: relaxation (a strategy aimed at the reduction of arousal), increase in leisure activities (behavioural-type strategy) and reduction of attention (cognitive-type strategy).

The studies in which functional-type strategies have been used present a somewhat confusing picture. In some of them it is concluded that patients diagnosed as psychotic basically use strategies aimed at problem-solving when coping with difficulties resulting from basic cognitive disorders and psychotic symptoms (Böker,

Figure 2
Structural coping strategies based on factoria analysis

O'SULLIVAN, 1994	CARTER, MAcKINNON & COPOLOV, 1996	FARHALL & GEHRKE, 1997
Hopeful and optimistic engagement	Strategies based on subvocal speech	Active acceptance
Despairing rejection		
	Search for competing auditory stimulation	Passive coping
Ambivalent acceptance		
	Well-integrated or intellectual responses to an intrusive stimulus	Resistance coping
Hopeful rejection		

Figure 3
Functional coping strategies

BÖKER, BRENNER ET AL. 1984 BRENNER, BÖKER ET AL. 1987 BÖKER, BRENNER ET AL. 1989	TAKAI, UEMATSU, KAIYA, INOUE & UEKI, 1990	WIEDL & SCHÖTTER, 1991 WIEDL, 1992	MUESER, VALENTINE & AGRESTA, 1997	MIDDELBOE & MORTENSEN, 1997
Compensatory efforts aimed at problem-solving	Distancing/avoidance	Coping level:	Cognitive coping	Coping level
				Cognitive control
	Behavioural change	Behavioural	Behavioural coping	Behavioural control
		Cognitive		Social change
Compensatory efforts not aimed at problem-solving	Strategic intervention	Emotional	Social coping	Physiological change
				Symptomatic behaviour
	Medical strategy	Direction of coping:	Non-social coping	
				Coping direction
	Resistance	- Problem-centred efforts		- Strategy aimed at problem-solving
		- Non-problem-centred efforts	Problem-focused coping	- Strategy not aimed at problem-solving
			Emotion-focused coping	

Brenner, Gerstner, Keller, Müller, & Spichtig, 1984; Brenner, Böker, Müller, Spichtig & Würgler, 1987; Takai, Uematsu, Kaiya, Inoue & Ueki, 1990; Middelboe & Mortensen, 1997); in others, it is demonstrated that they use emotion-oriented or non-problem-solving strategies (Van Den Bosch, Van Asma, Rombouts & Louwerens, 1992; Van Den Bosch & Rombouts, 1997; MacDonald et al., 1998); and there are also others in which none of these strategies predominates in such patients (Wiedl & Schötter, 1991; Wiedl, 1992; Mueser, Valentine & Agresta, 1997).

Given these difficulties, some authors have proposed the hypothesis that these differences can be explained by the mediation of other variables. Specifically, Wiedl & Schötter (1991), in an initial analysis of their research results, found no differences between frequency of use of emotion-oriented coping strategies and those aimed at problem-solving. However, in a second analysis, in which they grouped subjects according to their degree of subjective tension (high versus low), they found a clear response pattern that differed in accordance with this variable. Patients with high tension levels used mainly emotion-oriented coping strategies, whilst subjects with low tension used strategies aimed at problem-solving.

In this regard, Van Den Bosch & Rombouts (1997) found a relationship between cognitive variables and coping style. Their study distinguished three models of coping that correlate with certain patterns of cognitive functioning:

The first consists in the correlation between coping models based on problem-solving, distraction and cognitive acceptance (which they called “healthy coping”, and which corresponds in part to the coping strategies based on problem-solving) and self-reports of greater cognitive control (specifically the subjective report by these patients of greater processing capacity and greater attentional control).

The second model, called by these authors “demoralized coping” (which, as it can be seen, is similar to the construct of coping strategy based on emotion), consists in the correlation between coping strategies based on avoidance and worry, and dysfunctional cognitive functioning (overload and distractibility), the subjective experience of unease and a high level of mental effort while carrying out the cognitive performance tasks.

The third model is based on the correlation between coping strategies of the search for emotional support and expression and poor cognitive performance in objective attentional tasks (specifically in the Continuous Performance Test); that is, this dependent coping style is linked to poor objective processing skills without their being accompanied by subjective unease.

Finally, results from other studies provide similar data; for example, Pallanti, Quercioli & Pazzagli (1997) also demonstrated that level of subjective unease explains the differential use of coping strategies, while McDonald et al. (1998) showed that the different ratings of control of stressful situations on the part of schizophrenic patients determine the use of coping strategies based on problem-solving or on emotion, and this finding coincides with those of Wiedl & Schötter (1991) and Van Den Bosch & Rombouts (1997).

In sum, it appears that the use of a specific coping strategy may be determined by the individual’s own appreciation of his or her cognitive difficulties, and by the emotional burden experienced when faced with a given stressor. This means that strategies based on problem-solving are used by the most cognitively competent subjects, who therefore have less subjective tension, while those based on emotion are used by subjects with greater difficulties. This probably explains, on the one hand, the apparently contradictory results found in the literature with respect to functional strategies, and on the other, the massive use of strategies of a behavioural type, due to the lower cognitive cost to the subject involved in putting them into practice. Nevertheless, the correlational methodology employed in these studies does not permit us to determine whether these cognitive and emotional difficulties are the cause of certain coping styles or whether, on the other hand, certain coping styles or the nature and intensity of the stressors are the cause of these cognitive and emotional difficulties. This is a matter yet to be resolved empirically.

Relationship between coping strategies and symptoms presented by people diagnosed as psychotic

One group of studies has attempted to identify the relationship between psychotic and non-psychotic symptoms in schizophrenic patients and the use of coping strategies (Breier & Srauss, 1983; Cohen & Berk, 1985; Tarrier, 1987; Carr, 1988; Takai et al.,1990; Wiedl, 1992; Brazo, Dollfus & Petit, 1995; Carter et al., 1996; Middelboe & Mortensen, 1997; Mueser, Valentine & Agresta, 1997; McDonald et al.,1998).

Breier & Strauss (1983) and Tarrier (1987) found that psychotic patients do not use coping strategies aimed specifically at defending themselves against the undesirable effects of certain symptoms. Rather, they concluded that the subjects in their research use in an indiscriminate way those strategies that have proved most useful for them in the past.

While it is true that, in general, research shows that these patients use a wide variety of strategies for coping with the same symptom, in some works that use structural taxonomies it is found that cognitive strategies are preferentially used for coping with delusional ideas and hallucinations (Cohen & Berk, 1985; Carr, 1988; Wiedl, 1992; Boschi, Adams; Bromet, Lavelle, Everett & Galambos, 2000), and strategies of behavioural change for coping with anxiety, depression (Cohen & Berk, 1985; Carr, 1988), motor retardation and inhibition, thinking disorders (Carr, 1988) and some negative psychotic symptoms such as apathy (Mueser, Valentine & Agresta, 1997).

In studies that have used correlational-type methodology, negative associations have been found between coping strategies and general symptomatology. Takai et al. (1990), for example, found with regard to indices of general psychopathology that high scores in the total BPRS (Brief Psychiatric Rating Scale, by Overall & Gorham, 1962), and more specifically in the subscales of hostility, suspicion, thinking disorder and arousal, coincide with the use of few coping strategies based on behavioural change, distancing and avoidance, and with low total scores in general use of coping strategies.

With regard to the relationship between coping and positive and negative psychotic symptomatology, Middelboe & Mortensen, (1997) and McDonald et al. (1998) also found a negative correlation between the sum of coping strategies and strategies based on problem-solving, and negative symptoms measured by means of the SANS scale (Scale for the Assessment of Negative Symptoms, Andreasen, 1983); that is, the presence of high scores for negative psychotic symptoms coincides with the use of few coping strategies based on problem-solving, and with low scores in the total coping indices. These results are compatible with those of Wiedl (1992), who showed that emotion-oriented coping occurs much more frequently in those patients that present a higher quantity of negative symptoms. With regard to positive symptoms, Middelboe & Mortensen (1997), found a positive correlation between this type of symptom, measured with the SAPS scale (Scale for the Assessment of Positive Symptoms, Andreasen, 1984), and coping strategies based on emotion; that is, the presence of many positive symptoms also coincides with greater use of strategies based on emotion.

On considering the results of these studies, we once again come up against the problem of not knowing whether the different levels of symptomatology determine the greater or lesser use of coping strategies or whether, on the contrary, the greater use of coping strategies means that patients experience fewer symptoms of their disorder. In this regard, in an attempt to arrive at a somewhat more satisfactory approach to the solution of this problem, Middelboe & Mortensen (1997) applied a linear regression analysis to their data, and found that the total number of coping strategies used and strategies based on problem-solving are preceded by higher scores in general symptomatology according to the BPRS, low scores in negative psychotic symptoms according to the SANS and greater awareness of suffering from a mental disorder. Nevertheless, strategies based on emotion do not reveal such a model. Only if the BPRS was eliminated from the equation was it found that an increase in positive psychotic symptoms significantly predicted the use of coping strategies based on emotion.

Effectiveness of coping strategies in people diagnosed as psychotic

Although Lazarus & Folkman (1986) did not include in their definition of coping the component of the effectiveness of strategies for relieving, reducing, adapting or resolving patients’ difficulties, the great majority of authors working in the field of psychosis have considered this factor to be highly relevant, since it has fundamental implications at both the theoretical and technical level.

However, it is not an easy task to embark on the study of the effectiveness of coping, since, on analyzing the literature on the topic, we encounter two problems: one refers to the conceptualization of coping (already discussed at the beginning of this work), and therefore, what is understood by effectiveness from each of the perspectives; the second problem has to do with the methodology followed in its study. Below we shall describe each one of the different forms of understanding the effectiveness of coping and the methodology used for its assessment.

One approach considers that a coping strategy is effective for the simple fact that it is used by the patient, that is, *the most effective strategies are those that are most frequently used* (Cohen & Berk, 1985; Carr, 1988; O’Sullivan, 1994). From a quite similar point of view, other authors, such as Brazo, Dollfus & Petit (1995), state that when there is a significant relationship between a type of symptom and a strategy, we can conclude that it is significantly effective, and therefore more frequently used. In our view, while it is possible that in some cases strategies are used because they are effective, the mere fact of use is no guarantee of effectiveness. People may put into practice certain behaviours simply because they have not learned others in the course of their life, and not because these are more effective for coping with their difficulties. In fact, Carter et al. (1996) found that the coping strategies most frequently used by the subjects in their research were not the most effective, and that, indeed, the most effective ones were used only by quite a small proportion of patients.

A second approach deals with the problem of effectiveness by obtaining indirect measures from subjects, making inferences from other indicators. An example of such an approach is that of Falloon & Talbot (1981), who tried to relate coping mechanisms with patients' level of adaptation to the voices. They did so by means of a global clinical evaluation of each patient's adaptation to the hallucinatory experience using a three-point scale (good, normal and poor adaptation). In their study, effectiveness is assessed indirectly, with subjects grouped in these three categories in order to subsequently identify which coping mechanisms are most commonly used according to the adaptation levels. Thus, from this perspective, it is deduced that the most well-adapted patients (that is, those least affected or least incapacitated by their voices) are those that use the most effective coping strategies.

A comparable approach, though somewhat more complex, is that used by Lee, Lieh-Mak, Yu & Spinks (1993). Using a correlational-type methodology, they tried to discover the relationships between coping strategies and a series of result indicators, such as social or everyday life adjustment, quality of life and symptomatology.

Although these approaches may be considered as interesting attempts to deal with the problem in hand, they raise several unresolved issues. The first consists in that the concept of adaptation, or quality of life, is a wide and ambiguous one, and we cannot tell whether the subjects achieve better adaptation or quality of life because of certain coping strategies, or rather because of the intervention of other factors that also affect these variables (intensity of the hallucinatory experience, level of social support received, degree of chronicity of the disorder, subject's skills, treatment received, etc.). The other issue concerns the fact that we do not know the effectiveness of the strategy according to the patient's own, subject evaluation.

In a third type of approach, some authors have considered the problem of *effectiveness based on the patient's assessment* (e.g., Tarrier, 1987; Dittman & Schüttler, 1990; Carter et al., 1996; Middelboe & Mortensen, 1997; Mueser, Valentine & Agresta, 1997; MacDonald et al., 1998). In doing so, they have generally used ordinal, Likert-type measures, in which the subject must evaluate effectiveness on scales of three or more points. For example, Tarrier (1987) and Middelboe & Mortensen (1997) asked their patients to rate each strategy as "highly successful" (total disappearance of the symptoms), "moderately successful" (moderate reduction of the symptoms or temporary disappearance) or "scarcely or not at all successful". Another type of scale used are five-point analogical-digital ones, in which patients are asked about the success of or degree of satisfaction with the result of their coping strategies (Wiedl & Schötter, 1991; Mueser, Valentine & Agresta, 1997). It is important to note that all of these methodological approaches are characterized by defining effectiveness as a unidimensional construct, that is, *effectiveness is assessed solely in relation to success in the elimination of the problem*.

However, other works use a more complex conceptualization of the effectiveness of coping, though always within an orientation based on the elimination of the problem. It consists in considering it as a multidimensional construct. Farhall & Gehrke (1997), for example, asked the subjects of their study to rate coping strategies for auditory hallucinations according to the degree of control they could exercise over the voices, the reduction in anxiety levels and a global measure of coping effectiveness. In this regard, we found of great interest the multidimensional approach used in another work by Wahass and Kent (1997), also in relation to coping with auditory hallucinations. The dimensions assessed were as follows:

- Subject's capacity for eliminating the voices.
- Capacity for reducing anxiety caused by the voices.
- Capacity for ignoring them.
- Capacity for making the voices quieter or silent.
- Capacity for doubting their content and making them less credible.

Patients rated each dimension on a five-point scale, according to the strength of each of these capacities.

A final approach, though quite undeveloped, is that proposed by Romme & Escher (1989, 1996) and Romme, Honig, Noorthoorn & Escher (1992). As we pointed out at the beginning of the present work, the objective of coping from this perspective consists not in the elimination of the problem, as in the previous studies, but in the subject's adjustment to it. Thus, *effectiveness consists in evaluating subjects' capacity for integrating the hallucinatory experience, for example, into their daily life*, that is, achieving the acceptance of the voices as "part of themselves", not considering the experience as necessarily negative. To this end, the subjects of the study were grouped, according to the adjustment criterion described above, into good copers (those with or without psychiatric pathology for whom the voices were not a problem and who lived a normal life) and poor copers (those who had not achieved such integration of the symptoms in their daily life). On the basis of this classification, and employing the subjects' own reports, the authors studied the differences in the coping strategies.

Up to now we have discussed different ways of understanding the effectiveness of coping, and have given a quite general outline of the procedures for its assessment. At this point we should ask ourselves whether coping strategies are useful in patients diagnosed as psychotic. As the reader will understand, this is not an easy undertaking, given the polysemic nature of the concept and the fact that few studies have tried to *cope* in a direct way with the topic. We shall base our search for an answer to our question on all the studies carried out to date, except those that have employed the first approach we described, that which identifies effectiveness with the use of coping mechanisms. As we pointed out above, this a quite deficient way of understanding this concept.

In general, according to the results of these studies, it can be stated that the coping strategies used by psychotic patients yield effectiveness values ranging from moderate to low. Whilst only in the works of Tarrier (1987) and MacDonald et al. (1998) did subjects present a moderate capacity for eliminating or controlling their symptoms, in a larger group of studies (Wiedl & Schötter, 1991; Wiedl, 1992; Lee, Lieh-Mak, Yu & Spinks, 1993; Carter et al., 1996; Farhall & Gehrke, 1997; Middelboe & Mortensen, 1997) it was concluded that this capacity was low. And in another study (Takai et al., 1990) it was stated that the effectiveness of spontaneous coping strategies is not demonstrated. We found just two works that show that coping in psychotics is effective (Dittman & Schüttler, 1990; Mueser, Valentine & Agresta, 1997).

More specifically, where the negative results can be seen most clearly is in coping with auditory hallucinations: all the works reviewed showed that the majority of patients with a psychosis diagnosis use fairly ineffective resources for coping with the voices. If we consider psychotic symptomatology in general and negative symptoms in particular, the picture changes slightly, though the improvement is not a notable one. Dittman & Schüttler (1990) found that 86% of patients described specific changes in their behaviour due to the use of coping strategies, and Mueser, Valentine & Agresta (1997) found that they were effective for coping with negative symptoms such as apathy.

Detailed analysis of these studies does not provide us with sufficient data to conclude that coping strategies in themselves are ineffective; rather, and in line with the suggestion of Carter et al. (1996), all we can state is that patients use coping strategies that are of little use or ineffective. As can be seen from our review, there is in all the studies a percentage of subjects (albeit low) that are capable of resolving in an appropriate way the difficulties and problems related to their disorder. Thus, our next task is to decide which are the coping strategies that are successful in these subjects, and to identify the conditions of their application.

Although it may seem paradoxical, in the case of hallucinations, strategies of direct confrontation with the voices, based mainly on carrying out actions for their elimination (distraction, relaxation, humming, conversing with others, etc.), increase anxiety and the frequency of the voices. However, strategies based on acceptance of the symptom (e.g., listening attentively to the voices and accepting what they say) and on passive coping (doing nothing and depending rather on external sources of support, for example “putting my trust in God”) – that is, those that involve exposing oneself directly to the voices without resisting them, produce a reduction in anxiety, but not necessarily a reduction in the frequency of the voices (Farhall & Gehrke, 1997).

Romme & Escher (1989, 1996) and Romme, Honig, Noorthoorn & Escher (1992) found that subjects who cope well with the voices are those who have more capacity for ignoring them when they wish to, who use more selective listening and who put limits on them in a selective way – in sum, those who do not reject the voices, accepting them as just one more type of event in their lives; meanwhile, those who have difficulties for coping with them make more use of strategies based on distraction (doing sport, having a shower, watching television, meditation, yoga, etc.). Similar results are those obtained by Nayani & David (1996), who showed that attempts at control by their patients did not succeed in modifying the frequencies of the hallucinations, but that positive interaction with the voices (e.g., conversing with them or talking to others about them) reduced anxiety. On the other hand, these authors detected behaviours that increased the frequency of voices, such as watching television or listening to the radio, which are clearly distraction strategies.

In this regard, Brazo, Dollfus & Petit (1995), despite considering that the effectiveness of coping is related to its use, expressed their surprise on finding in their study that “there is not always a logical relationship between the existence of a coping strategy, its effectiveness, and its frequency of use: for example, acceptance of the voices is a little-used strategy, but it is highly effective” (p. 458).

With regard to the basic cognitive disorders of people diagnosed as schizophrenic and negative symptoms, there are data to show that coping strategies based on emotion are the most commonly used when the levels of subjective tension are very high, subjects showing a low level of satisfaction with the effectiveness of such strategies. However, in cases where the level of tension is lower, the coping strategies most frequently used are those based on problem-solving, and are rated by subjects as effective (Wiedl & Schötter, 1991; Wiedl, 1992 & MacDonald et al., 1998).

Thus, what we can deduce from these results is that people diagnosed as psychotic tend to use ineffective coping strategies when they are subject to very high levels of stress and tension. In these situations, which are probably the most usual in this type of patient, the only adaptive strategies are those based on acceptance of the disorder and non-confrontation. However, when stress levels are low, problem-solving strategies are the most commonly used, and also the most effective. This is probably due to the fact that application of the latter type of strategy requires conditions and cognitive effort that are impossible in high-stress situations.

An interesting result that we have found in many of the studies reviewed is that the effectiveness of the coping improves if subjects use several strategies at the same time, while, on the other hand, the ones who have difficulties in coping are those that use only a single strategy (Falloon & Talbot, 1981; Tarrier, 1987; Romme & Escher, 1989; Dittman & Schüttler, 1990; Wiedl & Schötter, 1991; Lee, Lieh-Mak, Yu & Spinks, 1993; Carter et al., 1996; Nayani & David, 1996; Middelboe & Mortensen, 1997; Mueser, Valentine & Agresta, 1997). Despite this consensus, there are some variations in the results obtained in these studies. While for some authors, such as Middelboe & Mortensen (1997), using several strategies constitutes a highly effective way of organizing coping, for Tarrier (1987) its effect is only moderate, and others, such as Nayani & David (1996), argue that the multi-strategy approach only has an influence in the reduction of anxiety associated with the symptoms. In any case, what emerges from these results is that perhaps what is relevant here is not the use of a specific strategy, but rather subjects' attitudes and their efforts to cope with the stressful events that arise in their everyday lives (Mueser, Valentine & Agresta, 1997).

Wiedl & Schötter (1991) & Wiedl (1992) found that the effectiveness of coping depends on subjects' *assessment* of the stressful event. The majority of patients consider the psychotic symptoms and basic cognitive disorders to be permanent, and not modifiable by they themselves (primary assessment, according to Lazarus & Folkman's theory). On the other hand, the large percentage of high perceived controllability (that is, the subject thinks s/he can modify, and therefore, control, the occurrence and intensity of the symptoms and basic disorders) shows that they believe they can influence the amount of stress on their own initiative (secondary assessment).

However, this last result appears to contradict somewhat their assessment of non-modifiability with regard to the stressful event and of their low satisfaction with the attempts at coping: how is it possible to assess an event as controllable if it is perceived as non-modifiable and one has low satisfaction with the results of coping? Wiedl & Schötter (1991) and Wiedl (1992) explain this contradiction by stating that either this assessment of controllability is unrealistic, or the criteria patients use for assessing the attempts at coping are inappropriate. Taking into account the results of the work of Nayani & David (1996) and Farhall & Gehrke (1997), Wiedl's findings of subjects that present a pattern of low perceived modifiability, high perceived controllability and low satisfaction with coping for symptoms and basic disorders are not contradictory; quite the opposite, in fact, since we have seen that attempts to control symptoms (especially of hallucinations) seem not to eliminate the problem, but rather to exacerbate it.

Finally, an aspect worth discussing is the greater effectiveness of coping when the patient is capable of detecting the antecedents or situations that elicit the symptoms (Talbot & Falloon, 1981; Breier & Strauss, 1983; Tarrier, 1987; Thurm & Haefner, 1987; Brazo, Dollfus & Petit, 1995; MacDonald et al., 1998). Notable among these antecedents are, for example, internal tension, insomnia or nightfall in the case of hallucinations (Brazo, Dollfus & Petit, 1995). In the case of psychotic symptoms in general, Thurm & Haefner (1987) distinguished two groups of antecedents, which they called socio-emotional (e.g., conflicts with friends or relatives, or intense emotions in close social relationships) and socio-cognitive (e.g., psychological and physical tension, disorders of life rhythms, or complex social interactions).

Interest in this topic stems from theoretical models of self-control, such as that developed by Breier & Strauss (1983) in individuals diagnosed as psychotic, and in which it is postulated that the coping process comprises three phases: detection of antecedents, assessment of them as dangerous or threatening, and finally, the use of self-control behaviour.

Talbot & Falloon (1981) were the first to point out the importance of this variable. They discovered that one of the most important differences between good and poor copers was that the former tended to have a clear understanding of the antecedents associated with the onset of symptoms, so that they were able to easily avoid those situations that elicited them. Nevertheless, Tarrier (1987) did not fully support these conclusions: while he considered it relevant that subjects detected such antecedents, he also argued that this condition was not in itself sufficient if subjects had not previously learned coping skills. In any case, it seems clear to some authors that the success of self-control and coping mechanisms depends on subjects being aware that they are suffering from a disorder and, therefore, of the circumstances that improve it or worsen it (Dittmann and Schüttler, 1990; Takai and cols., 1990; Nayani and David, 1996; Middelboe and Mortensen, 1997).

CONCLUSIONS

The review of relevant studies we have presented here highlights the fact that there is no single and consensus-based concept of the term coping in the field of psychotic disorders, but rather different conceptions and uses of it. These differences are related to two basic positions: one of these is the approach that understands coping as behaviour whose objective is purely defensive (e.g., Tarrier, 1987), whilst in the other approach, the objective is adaptation and integration (e.g., Romme & Escher, 1989). The second idea has to do with the assessment of its effectiveness. As we have seen, for many of the authors mentioned here it is important to understand coping according to its effects on the patient's problems, whilst for other authors, patients' efforts for managing stressful demands are more important than the results themselves (Lazarus & Folkman, 1986).

People diagnosed as psychotic are not passive in the face of difficulties related to their illness. They *claim* to put into practice a series of behaviours with the objective, in accordance with the different forms of understanding coping, of eliminating these problems or adapting themselves to them. Nevertheless, it should be stressed that in the studies reviewed this information has always been obtained from the patient's retrospective report in an interview situation with the researcher, and never from the reports of others, such as family, carers or trained observers.

Although the coping strategies used by psychotic patients are multiple and varied, we have seen in this work that they can be grouped in two large categories: structural and functional. The combined use of these two categories for the description and study of coping can provide more information and enrich our understanding of the different forms in which people with this disorder relate to their illness. In this regard, various authors have used definitions based on two dimensions: *coping level*, which refers to the strategies we have called structural, and *coping direction*, which would include the functional taxonomies (e.g., Wiedl and Schötter, 1991; Wiedl, 1992).

The strategies most commonly used from a structural point of view are behavioural ones. However, from a functional point of view the results are not clear. It would appear that coping styles are mediated by other variables, such as subjective tension and cognitive variables (e.g., Wiedl and Schötter, 1991; Van Den Bosch and Rombouts, 1997). When subjects have high levels of tension and difficulties related to attentional level and information processing, we find that they more frequently use strategies of a behavioural type and oriented towards emotion. When tension is low and cognitive competence high, it is more common to find efforts aimed at problem-solving.

These behaviours may be used to cope with both general and specifically psychotic symptomatology. In this regard, the results of research reveal that psychotic individuals tend to use spontaneously strategies of a cognitive nature to cope with florid psychotic symptoms, such as hallucinations and delusions, and behavioural-type strategies for negative symptoms and others that are not specifically psychotic. Also, it seems that high levels of positive and negative psychotic symptomatology are related to less use of coping strategies aimed at problem-solving and greater use of emotion-oriented strategies.

In general, the coping strategies spontaneously used by psychotic patients are ineffective in a large percentage of subjects. In any case, there is a minority that do present higher levels of satisfaction in coping with the disorders caused by their illness. These subjects use different mechanisms according to the type of problems they face. When they have high levels of stress or helplessness, as may occur with hallucinations or positive symptoms in general, the most effective strategies are those based on acceptance of the disorder

As we saw previously, attempting to cope directly with the symptom by means of self-control or problem-solving techniques (as in the case of strategies aimed at problem-solving), or through negation or avoidance (as in emotion-oriented strategies), leads not to its elimination, but rather quite the contrary, its consolidation, and an increase in the associated anxiety. Clearly, when the conditions of subjective stress or tension are high, or the cognitive deficits serious, the subject can only use those strategies that involve low cognitive cost, such as those based on emotion. However, subjects that do not fight the symptoms, and do not attempt to avoid them either, but rather to accept them, succeed in reducing the anxiety associated with the symptoms and living more comfortably, even without managing to make them disappear. In this context, and based on the analysis of studies on coping, we understand as *acceptance* the direct experience of the problems associated with psychotic symptoms without defending oneself from them – that is, entering into contact with the thoughts, emotions and situations related to the symptoms without attempting to reduce or eliminate them, but on the contrary, adapting oneself to them and incorporating them as just another event in one's life.

Nevertheless, the situation changes completely when people are subject to low levels of stress, or when the type of symptomatology does not cause feelings of helplessness. In such situations it is easier to put into practice self-control and problem-solving strategies, with higher levels of satisfaction being found for the results obtained. This tends to occur, for example, in the case of coping with non-psychotic symptomatology (anxiety and depression), low-intensity social relationships, inactivity and feelings of apathy.

The variables we have found in the literature that can increase the effectiveness of coping and one's satisfaction with it are basically: the use of multiple strategies for coping with the same problem, assessments of modifiability of a stressful event (primary assessment) and controllability of that event with one's own resources (secondary assessment), and subject's level of awareness of the problem's antecedents and of suffering from a disorder.

As the reader might suppose, the above conclusions are merely provisional, since many of these results need to be replicated in order to confirm the validity of the interpretations we have made of the data provided by the studies. Moreover, it is quite lamentable to observe that, although we came across a considerable number of articles that dealt with this topic, the majority of them were carried out with poor methodology (basically descriptive, and in few cases of a correlational type), and described even by their own authors as no more than exploratory. This demonstrates the scarce tradition and continuity of research on coping in psychotics, despite its high theoretical relevance. It would appear that the interest may be in simply demonstrating that people diagnosed as psychotic are capable of coping with their disorders, in order to justify the subsequent development of intervention programs based on the enhancement of spontaneous coping strategies (e.g., Tarrier, Beckett, Harwood, Baker, Yusupoff, Ugarteburu, 1993).

This assumption is supported, on the one hand, in the fact that we have found no stable research lines on the topic (except in some researchers of German origin, such as Wiedl, Böker or Brenner), but rather isolated works with scant continuity among them, and on the other, in the fact that the research problems considered are almost always the same ones, with no serious conceptual and theoretical reflection on the topics and variables studied, and with no connection to research on coping in other populations and disorders.

Furthermore, and continuing with the argument that there is a lack of coherence and continuity in research on coping, we find that the application of the knowledge obtained is being carried out directly, with no review of the totality of results yielded in the field. Let us take as an example the case of the strategies developed by Tarrier et al. for training psychotic patients to use their own coping skills effectively (referred to in the literature as Coping Strategy Enhancement, CSE). Tarrier states that through CSE techniques it is attempted to identify the coping strategies patients use naturally, and then to teach them to use in a systematic way these strategies and other new ones, so that they have access to a wide repertoire of ways of coping with their symptoms (Tarrier et al., 1993).

However, we have demonstrated in this work that many of the strategies trained in these programmes are fairly ineffective, an example being the use of distraction techniques or direct confrontation by means of self-instructions; and moreover, they do not take into account factors such as subjects' tension level or cognitive deficits, which, as we saw above, condition the type of strategies to be used (acceptance versus problem-solving). A test of all this, as we concluded in a previous work (Perona Garcelán & Cuevas Yust, 1999), is that these strategies are not effective in the control or elimination of auditory hallucinations, have failed to demonstrate their superiority over other psychological treatments and do not have lasting effects.

Even so, and despite what has been said up to now, we consider that the training of psychotic patients in coping strategies could be included in any treatment package, as long as, in order to enhance its effectiveness, the following aspects are taken into account:

1. To evaluate how the patient's subjective tension and cognitive abilities influence the use of coping behaviours.
2. To bear in mind the primary and secondary assessments with respect to each of the symptoms considered.
3. When the subjective tension level is quite high, to be sure to use strategies based on acceptance, and when it is moderate or low, strategies based on problem-solving.
4. To avoid teaching patients to use strategies based on emotion, except when their level of deterioration is very high and they present a marked state of helplessness.
5. To consider the use of multiple strategies to deal with a single problem.
6. To teach the subject to identify the antecedents related to the problem in question.
7. To teach and assess the results of the strategies in a multidimensional way that is not centred on elimination of the disorder.

By way of conclusion, it can be stated that, on the basis of data provided by research to date, coping strategies in psychotics constitute a highly relevant variable for the understanding of the psychological factors involved in psychotic symptomatology. In fact, we have seen how certain coping styles can cause a patient's anxiety or hallucinations to increase or, on the contrary, to decrease. This leads us to suggest that theoretical models on the etiology of psychotic symptoms should take into account that the way of reacting to or coping with symptoms and daily life problems will influence the mechanisms that contribute to the onset, maintenance and possible disappearance of these symptoms. For these reasons, we consider it necessary to foment the study of coping in psychotics, and not to undervalue it, as do some authors of a cognitive orientation (Chadwick & Birchwood, 1994). In the near future, work in this field will probably permit us to widen our knowledge of psychosis and the means of effectively "coping" with it.

REFERENCES

Andreasen, N.C. (1983). *The Scale for the Assessment of Negative Symptoms* (SANS). Iowa City: University of Iowa.

Andreasen, N.C. (1984).*The Scale for the Assessment of Positive Symptoms* (SAPS). Iowa City: University of Iowa.

Böker, W.; Brenner, H.D.; Gerstner, G.; Keller, F.; Müller, J. & Spichtig, L. (1984). Self-healing strategies among schizophrenics: attempts at compensation for basic disorders. *Acta Psychiatrica Scandinavica, 69*, 373-378.

Böker, W.; Brenner, H.D. & Würgler, S. (1989). Vulnerability-linked deficiencies, psychopathology and coping behaviour of schizophrenics and their relatives. *British Journal of Psychiatry, 155,* (supplement 5),128-135.

Boschi, S.; Adams, R.E.; Bromet, E.J.; Lavelle, J.E.; Everett, E. & Galambos, N. (2000). Coping with psychotic symptoms in the early phases of schizophrenia. *American Journal of Orthopsychiatry, 70(2),* 242-252.

Brazo, P.; Dollfus, S. & Petit, M. (1995). Stratégies anti-hallucinatoires expérimentées par les schizophrènes. *Annales Medico Psychologiques, 153(7),* 456-459.

Breier, A. & Strauss, J.S. (1983). Self-control in psychotic disorders. *Archives of General Psychiatry, 40,* 1141-1145.

Brenner, H.D.; Böker, W.; Müller, J.; Spichtig, L. & Würgler, S. (1987). On autoprotective efforts of schizophrenics, neurotics and controls. *Acta Psychiatrica Scandinavica, 75*, 405-414.

Carr, V. (1988). Patients' techniques for coping with schizophrenia: an exploratory study. *British Journal of Medical Psychology, 61*, 339-352.

Carter, D.M.; Mackinnon, A. & Copolov, D.L. (1996). Patients' strategies for coping with auditory hallucinations. *The Journal of Nervous and Mental Disease, 184 (3),* 159-164.

Chadwick, P. & Birchwood, M. (1984). The omnipotence of voices. A cognitive approach to auditory hallucinations. *British Journal of Psychiatry, 164*, 190-201.

Cohen, C.I. & Berk, L.A. (1985). Personal coping style of schizophrenic outpatients. *Hospital and Community Psychiatry, 36(4)*, 407-410.

Crespo, M. & Cruzado, J.A. (1997). La evaluación del afrontamiento: Adaptación española del cuestionario COPE con una muestra de pacientes universitarios. *Análisis y Modificación de Conducta, 23(92),* 797-830.

Dittman, J. & Schüttler, R. (1990). Disease consciousness and coping strategies of patients with schizophrenic psychosis. *Acta Psychiatrica Scandinavica, 82*, 318-322.

Falloon, I.R.H. & Talbot, R.E. (1981). Persistent auditory hallucinations: coping mechanisms and applications for management. *Psychological Medicine, 11*, 329-339.

Farhall, J. & Gehrke, M. (1997). Coping with hallucinations: exploring stress and coping framework. *British Journal of Clinical Psychology, 36*, 259-261.

Folkman, S. & Lazarus, R.S. (1980). An analysis of coping in a middleaged community sample. *Journal of Health and Social Behavior, 21*, 219-239.

Frederick, J. & Cotanch, P. (1995). Self-help techniques for auditory hallucinations in schizophrenia. *Issues in Mental Health Nursing, 16*, 213-224.

Galán Rodríguez, A. (2000). *Conducta de enfermedad, estrategias de afrontamiento y calidad de vida.* Unpublished doctoral thesis. Universidad de Sevilla.

Kanas, N. & Barr, M.A. (1984). Self-control of psychotic productions in schizophrenics. *Archives of General Psychiatry, 41*, 919-920.

Kinoshita, F.; Yagi, G.; Inomata, T. & Kanba, S. (1991). Coping style of schizophrenic patients in the recovery from acute psychotic state: a preliminary study. *Keio Journal of Medicine, 40(3),* 129-131.

Kumar, S.; Thara, R. & Rajkumar, S. (1989). Coping with symptoms of relapse in schizophrenia. *European Archives of Psychiatry and Neurological Sciences,239*, 213-215.

Lange, H.U. (1981). Anpassungsstrategien, bewältigungsreaktionen und selbstheilungsversuche bei schizophrenen. *Fortschr Neurol Psychiatr, 49,* 275-285.

Lazarus, R. S. & Folkman, S. (1986). *Estrés y procesos cognitivos.* Barcelona: Martínez Roca (original edition, 1984).

Lee, P.W.H.; Lieh-Mak, F.; Yu, K.K. & Spinks, J.A.

(1993). Coping strategies of schizophrenic patients and their relationship to outcome. *British Journal of Psychiatry, 163,* 177-182.

López-Roig, S.(1991). *Determinantes psicosociales del estrés y su afrontamiento en pacientes quirúrgicos*. Unpublished doctoral thesis. Universidad de Alicante.

MacDonald, E.M.; Pica, S.; McDonald, S.; Hayes, R.L. & Baglioni, A.J. (1998). Stress and coping in early psychosis. *British Journal of Psychiatry, 172,* (supplement 33),122-127.

McNally, S.E. & Goldberg, J.O. (1997). Natural cognitive coping strategies in schizophrenia. *British Journal of Medical Psychology, 70,* 159-167.

Middelboe, T. & Mortensen, E.L. (1997). Coping strategies among the long-term mentally ill: categorization and clinical determinants. *Acta Psychiatrica Scandinavica, 96*, 188-194.

Mueser, K.T.; Valentine, D.P. & Agresta, J. (1997). Coping with negative symptoms of schizophrenia: patient and family perspectives. *Schizophrenia Bulletin, 23(2),* 329-339.

Nayani, T.H. & David, A.S. (1996). The auditory hallucination: a phenomenological survey. *Psychological Medicine, 26*, 177-189.

Nuechterlein, K.H. & Dawson, M.E. (1984). A heuristic vulnerability/stress model of schizophrenic episodes. *Schizophrenia Bulletin, 10*, 300-312.

O'Sullivan, K. (1994). Dimensions of coping with auditory hallucinations. *Journal of Mental Health, 3,* 351-361.

Overall, J.E. & Gorham, D.R. (1962). The brief psychiatric rating scale. *Psychological Report, 10,* 799-812.

Pallanti, S.; Quercioli, L. & Pazzagli, A. (1997). Relapse in young paranoid schizophrenic patients: a prospective study of stressful life events, P300 measures, and coping. *American Journal of Psychiatry, 154(6),* 792-798.

Perona Garcelán, S. & Cuevas Yust, C. (1996). Intervenciones cognitivo conductuales sobre las alucinaciones auditivas en sujetos psicóticos. Una revisión. *Psicologemas 10(20),* 225-256.

Perona Garcelán, S. & Cuevas Yust, C. (1999). Efectividad de la terapia cognitivo-conductual individual aplicada a los síntomas psicóticos. 2. Estudios aleatorizados con grupos control. *Apuntes de Psicología, 17(3),* 249-274.

Romme, M.A.J. & Escher, A.D.M.A.C. (1989). Hearing voices. *Schizophrenia Bulletin, 15(2),* 209-216.

Romme, M.A.J. & Escher, A.D.M.A.C. (1996). Empowering people who hear voices. In G. Haddock & P.D. Slade: *Cognitive-behavioural interventions with psychotic disorders.* London: Routledge.

Romme, M.A.J.; Honig, A.; Noorthoorn, E.O. & Escher, A.D.M.A.C. (1992). Coping with hearing voices: an emancipatory approach. *British Journal of Psychiatry, 161*, 99-103.

Takai, A.; Uematsu, M.; Kaiya, H.; Inoue, M. & Ueki, H. (1990). Coping styles to basic disorders among schizophrenics. *Acta Psychiatrica Scandinavica, 82*, 289-294.

Tarrier, N. (1987). An investigation of residual psychotic symptoms in discharged schizophrenic patients. *British Journal of Clinical Psychology, 26*, 141-143.

Tarrier, N.; Beckett, R.; Harwood, S.; Baker, A.; Yusupoff, L.; Ugarteburu, I. (1993). A trial of two cognitive-behavioural methods of treating drug-resistant residual psychotic symptoms in schizophrenic patients: I. Outcome. *British Journal of Psychiatry, 162*, 524-532.

Thurm, I. & Haefner, H. (1987). Perceived vulnerability, relapse risk and coping in schizophrenia. An exploratory study. *European Archives of Psychiatry and Neurological Sciences, 237,* 46-53.

Van Den Bosch, R.J. & Rombouts, R.P. (1997). Coping and cognition in schizophrenia and depression. *Comprehensive Psychiatry, 38(6),* 341-344.

Van Den Bosch, R.J.; Van Asma, M.J.O.; Rombouts, R.P. & Louwerens, J.W. (1992). Coping style and cognitive dysfunction in schizophrenic patients. *British Journal of Psychiatry, 161,* (supplement 18), 123-128.

Wahass S. & Kent, G. (1997). Coping with auditory hallucinations: a cross-cultural comparison between Western (British) and Non-Western (Saudi Arabian) patients. *The Journal of Nervous and Mental Disease, 185(11),* 664-668.

Wiedl, K.H. (1992). Assessment of coping with schizophrenia. Stressors, appraisals, and coping behaviour. *British Journal of Psychiatry, 161*, (supplement 18), 114-122.

Wiedl, K.H. & Schötter, B. (1991). Coping with symptoms related of schizophrenia. *Schizophrenia Bulletin, 17(3),* 525-538.

Yagi, G.; Kinoshita, F. & Kanba, S. (1992). Coping style of schizophrenic patients in the recovery from acute psychotic state. *Schizophrenia Research, 6,* 87-88.

Yusupoff, L. & Tarrier, N. (1996). Coping strategy enhancement for persistent hallucinations and delusions. In G. Haddock & P.D. Slade: *Cognitive-behavioural interventions with psychotic disorders*. London: Routledge.

Zubin, J. & Spring, B. (1977). Vulnerability – a new view of schizophrenia. *Journal of Abnormal Psychology, 86*, 103-126.

Copyright 2002 by the
Colegio Oficial de Psicólogos. Spain

Psychology in Spain, 2002, Vol. 6. No 1, 41-55

TEAMWORK IN DIFFERENT COMMUNICATION CONTEXTS: A LONGITUDINAL STUDY

Ana Zornoza Abad, Virginia Orengo Castellá, Isabel Gosálvez Cuenca and Pilar González Navarro
University of Valencia

The increasing application of new technologies to work teams in organizational contexts makes necessary a careful study of both work group processes and results. In this context, previous research has provided evidence on how mediated communication and task type have a differential impact on group functioning and results (coordination of discussion, participation, influence of dominant members and normative control). However, in spite of advances in this area, there has been a lack of studies examining work group functioning when teams are permanent, probably because of the high cost of longitudinal studies. Temporal aspects do, however, play an important role in group processes and their results. From this perspective, the objective of this study is to analyze how group processes develop as they adapt to the different communication media and as a function of time and task type. Data were collected from a laboratory study in which 31 groups of 4 members participated. Groups were randomly assigned to different communication conditions (face-to-face with computer support, videoconference and e-mail), all groups performing different types of task (creativity, intellective and conflict) and meeting on several occasions over eight work sessions.

En la actualidad, la creciente implantación de nuevas tecnologías de la información para el trabajo en grupo en los contextos laborales hace necesario profundizar en el estudio de los procesos y resultados del trabajo en grupo. En este sentido, la investigación realizada hasta el momento ha puesto de manifiesto el impacto diferencial de la comunicación mediada y del tipo de tarea utilizado tanto en aspectos del funcionamiento grupal como de resultados. A pesar de los avances en este ámbito, se observa una escasez de estudios que exploren el funcionamiento de los grupos de trabajo cuando éstos poseen un carácter permanente, probablemente por los elevados costes que supone llevar a cabo estudios longitudinales. En este sentido los aspectos temporales juegan un importante papel en las conductas mostradas por los miembros durante la interacción grupal y en sus resultados. Desde esta perspectiva, el objetivo del presente trabajo consiste en analizar cómo varía el funcionamiento de los grupos al adaptarse éstos a los distintos medios de comunicación a lo largo del tiempo y en función del tipo de tarea utilizado. Para ello, se ha realizado un experimento de laboratorio en el que han participado 31 grupos de cuatro miembros cada uno. Los grupos han sido asignados aleatoriamente a las diferentes condiciones de comunicación (cara a cara con apoyo electrónico, videoconferencia y correo electrónico). Todos los grupos realizaban distintos tipos de tareas (creatividad, intelectivas y de conflicto) y se reunían en varias ocasiones durante ocho sesiones de trabajo consecutivas.

As we enter the twenty-first century, there would appear to be a consensus among researchers in affirming that one of the most distinctive characteristics of the twentieth century was a high level of technological development and its progressive application to all areas of society. The emergence of automated systems has had an enormous influence on our society, which is even referred to using terms such as the *post-industrial society*, *cybersociety* or *technotronic society* (Nelson, 1995).

Technological advances have been applied to diverse areas of social and working life in our society. Notable among current tendencies in work organizations is a substantial change in their structure and functioning. The need for flexibility in order to adapt to rapid changes in the socio-labour context means that organizations tend to adopt “networked” forms, in which there are neither internal limits nor hierarchical differentiation, but rather a predominance of cooperative relationships and flexible patterns of communication, with work teams becoming the basic structural units (DeSanctis and Poole, 1997; Alcóver and Gil, in press). At the same time, new, computer-based communication systems make possible these new forms of organization and

---

The original Spanish version of this paper has been previously published in *Apuntes de Psicología*, 2001, Vol. 19, No 1, 137-160

....

Correspondence concerning this article should be addressed to Ana Zornoza, UIPOT: Unidad de Investigación de Psicología del Trabajo y de las Organizaciones, Universidad de Valencia, Area de Psicología Social, Facultad de Psicología, Avda Blasco Ibañez, 21, 46010 València. Spain. Ana.Zornoza@uv.es

influence teamwork by allowing the automatization of the distribution of information among group members and the structuring of group discussion, and by providing problem-solving techniques such as decision modelling or patterns for the planning of group activity (Johansen, 1988; Huber, 1990). It is, therefore, clearly interesting and necessary to study, from a psychosocial perspective, the effect of the implementation and use of these new information technologies on teamwork in the organizational context.

The study of groups within the framework of Social and Organizational Psychology has led to the development of various theoretical models that attempt to identify critical factors affecting the functioning and effectiveness of work groups (structural, ecological and motivational models) (Hackman and Morris 1975; Hackman 1987; Sundstrom, De Meuse and Futrell, 1990 and Guzzo, Yost, Campbell and Shea, 1993). All of these aim to analyze and explain group functioning on the basis of different types of variables (contextual, process and result). Moreover, they share the idea that group effectiveness is explained not solely by the final result obtained by the group, but also by the process followed in order to arrive at that result. McGrath (1984) defines group processes as “the behaviour pattern of its members”. This refers to the patterns of communication that emerge between the group members, interpersonal actions, influences and contacts and the active use of available technologies. Interaction processes mediate the relationship between inputs (technology, task, members’ characteristics) and results.

However, many of these models were developed for a traditional, face-to-face communication situation, so that it would be necessary to adapt them or to develop new models of functioning for groups working with new information technologies. From the 1970s, research teams in various countries developed with the aim of analyzing the influence of computer-mediated communication on group work from a psychosocial perspective (Peiró, Prieto and Zornoza, 1993; McGrath and Hollingshead, 1994).

The main findings obtained in this field indicate that group processes differ according to the technology used in the interaction. Numerous studies have found that groups that communicate electronically focus more on activities related to the task than groups that communicate face-to-face, invest more time in completing the task and have more problems to reach a consensus, and that group members participate to an equal extent (Kiesler, Siegel and McGuire 1984; Siegel, Dubrovsky, Kiesler and McGuire 1986; Kiesler and Sproull, 1992; Hiltz et al. 1986). These have been considered as advantages of mediated communication, with groups being more efficient on investing less time in non-task-related interpersonal activities (Steiner, 1972). More negative socioemotional behaviour has been found in mediated communication, whilst it is more positive in rich media, such as face-to-face communication or videoconference (Smolensky, Carmody and Halcomb, 1990; Siegel et al. 1986; Dubrovsky, Kiesler and Sethna, 1991). On the other hand, some authors, far from finding a higher percentage of negative interpersonal behaviour in mediated communication (Spears and Lea, 1992; Strauss, 1997), have found that these groups even develop more prosocial behaviour than those that communicate face-to-face (Walther, 1995). As regards the level of conflict found, there are also two alternative positions. On the one hand are studies that find higher levels of conflict in mediated communication (Kiesler et al. 1984), due to the impossibility of transmitting non-verbal signals that regulate conflictive behaviour, on the other, there are those that state that groups whose interaction is mediated by new technologies experience lower levels of conflict than those that communicate face-to-face because they pay more attention to the task and less to interpersonal relationships (O’Connor, Gruenfeld and McGrath, 1993).

Thus, we can group the research carried out in this field according to two different theoretical approaches. The first approach is based on theories of “medium-richness” (Short, Williams and Christie, 1976; Daft and Lengel, 1986), and argues that it is the objective characteristics of the communication medium that determine group functioning. The less rich a medium, the more “cue filtering” there is, so that non-verbal or social cues cannot be transmitted during group interaction. This exclusivity attributed to technology has recently been questioned by several researchers, giving rise to a new theoretical perspective. This alternative approach is based on the theory of social information processing and on Social Identity theory, defining the group as an open and multifunctional system (Spears and Lea, 1992; Walther, 1994; 1996). These authors point out that technology can influence group functioning positively or negatively, since changes in interaction processes do not depend solely on the technology itself; rather, there are other relevant variables that affect this relationship, such as the structure of the group before using the technology, type of task or the time spent using the medium.

This approach therefore permits the adoption of a new concept of group based on a multifunctional system. The group has multiple functions that are of equal importance for its effectiveness, not only that of production, but also in relation to group well-being and the personal development of its members (McGrath, 1990).

Furthermore, the technology interacts with the type of task the group has to carry out (McGrath and Hollingshead, 1994; Strauss and McGrath, 1994), so that the influence of the technology on group processes and results will depend on the fit between the richness of the communication medium and the demands of the group's task. A task involving the generation of ideas does not demand the same level of coordination and interdependence between group members as an intellective task or a mixed task in which group members show a conflict of interests, ideas or values (Argote and McGrath, 1993). Also, the fit between task and technology is not static, but dynamic, changing over time (McGrath and Berdahl, 1998). That is, as time passes, the group acquires experience in the use of the medium, so that the group adapts to its objective characteristics and develops new strategies for carrying out the task, which may mean that it accomplishes the task effectively with a less rich medium.

Thus, McGrath proposes that we consider the interaction of three elements: technology, task and time, and it is on the basis of these that he develops his model of group functioning (Time, Interaction and Performance, TIP, 1990, 1991), on which our study will be based. In this model, McGrath stresses the importance of time, starting out from the idea that groups working in organizations carry out activities that involve a certain temporal continuity, which makes it necessary to carry out longitudinal studies. Likewise, and in accordance with the conception of the group as a multifunctional social system, this author distinguishes three group functions: production, group well-being and personal development of group members. In order to develop these functions the group passes through a series of stages that are neither fixed nor sequential (McGrath and O'Connor, 1996). These stages are choice of goals, solution of problems, resolution of conflicts and execution. Thus, in each stage and for each function the process of group interaction may present different patterns.

Andriessen and Van der Velden (1993) propose another model of group interaction, adapted from McGrath's (1990), in which they distinguish two functions of group interaction: one oriented to the task and another oriented to the socioemotional maintenance of the group. Furthermore, this model suggests that on performing the task the group will pass through two stages: that of preparation and that of execution.

From this longitudinal perspective, the empirical evidence obtained indicates that group interaction processes will differ according to the group's stage of development. In this line, Lebie, Jonathan and McGrath (1996) found that groups tend, over time, to decrease their communication in group processes related to the task (planning, composition and procedure), and that this is not the case in group processes of a socioemotional nature. That is, as the members adapt to the group and the task, they need to focus their energy less on the function of production and to be more oriented towards the group – to fulfil the functions of group well-being and support for members.

Andriessen and Van der Velden (1993) also argue that preparation activities will be more relevant during the initial stages of a group's life. After this initial stage groups tend to develop more or less fixed patterns of interaction, and invest less and less time in developing strategies or patterns of interpersonal relationships. These authors also found that the conventional interaction condition (face-to-face) facilitates, to a greater extent, the function of socioemotional maintenance of the group, in comparison to mediated interaction conditions. Moreover, the results obtained by Walther and Burgoon (1992) over time show that groups that interact by means of mediated communication focus less and less on the task and more on socioemotional relationships.

Thus, in general terms, the results of research in this field appear to indicate that groups using mediated communication need more time to adapt to the technology used and develop new strategies that permit them to be effective. This may initially condition the stability of their interaction processes and limit their performance (McGrath, 1993; Mennecke, Hoffer and Wynne, 1992). However, with the passage of time and continued use of the technology, group processes become established and differences that depend on the communication medium disappear.

PREVIOUS STUDIES BY THE UIPOT ON GROUP INTERACTION PROCESSES IN MEDIATED COMMUNICATION CONTEXTS

The Work and Organizational Psychology Research Unit (*Unidad de Investigation de Psicología del Trabajo and de las Organizaciones*, UIPOT) at the University of

Valencia has extensive experience in the study of the influence of new information technologies on group work. Since 1989 this research team has developed a sequence of three projects, which, despite having similar objectives, differ as regards the complexity of the approach and design of the study in each case, which increases as a result of the findings obtained both by our own team and by other researchers.

The first two research projects focused on the analysis of the psychosocial factors involved in the structure, functioning and effectiveness of cooperative work in groups that communicated by means of new technologies to resolve a task, and of the influence of the technology on group processes and results as a function of the type of task carried out by the group. Both projects used two interaction technologies as alternatives to conventional face-to-face interaction: videoconference and e-mail. Both studies were carried out in the laboratory with samples consisting mainly of university students (though the sample in the first study also included teachers from the Psychology Faculty and professionals and workers from different sectors), and both were transversal in design (Zornoza, 1992; Orengo, 1994; Acín, 1995). The principal difference between the two studies resided in the inclusion in the second one of different types of group tasks. Specifically, and based on the group task model proposed by McGrath (1984; Argote and McGrath, 1993), three types of task were considered: idea generation tasks, intellective tasks and value conflict tasks. The general objective of this second project consisted in analyzing the influence of the communication channel on group processes and results according to the type of task carried out.

In general, it can be stated that the theoretical approach followed by the research team was that which was predominant at that time, and which proceeded from the tradition defended by the theories of Social Presence (Short et al., 1976) and Medium-Richness (Daft, Lengel and Trevino, 1987). These theories used the argument of the filtering of social and non-verbal cues to explain the influence of the technology on group processes and results. However, in our studies we consider an additional variable that may influence the relationship between the technology and group functioning: *perception of the communication medium* (Peiró, Prieto and Zornoza, 1994; Zornoza, Ripoll and González, 1995).

The results obtained in our studies confirm, in a general way, the predictions made on the basis of this theoretical orientation. Groups that perform intellective tasks in rich media such as face-to-face or videoconference coordinate themselves better, develop decision strategies based on consensus, make more effort to accomplish the task, manage conflicts that emerge between group members in a positive way and present more positive socioemotional behaviour than when they interact via e-mail (Orengo, 1994; Acín, 1995; Orengo, Gosálvez, Fernández and Prieto, 1995; Orengo, Zornoza, Acín, Prieto and Peiró, 1996; Peiró, Prieto, Zornoza and Ripoll, 1999). The interaction between communication medium and type of task shows that in idea generation tasks there were no significant differences in the group interaction processes. Also, nor was there more negative socioemotional behaviour in the e-mail condition in comparison to videoconference and face-to-face in conflict of values tasks (Marzo, Rodríguez, Ripoll and Marín, 1995).

The third research project carried out in this line of work had as its main objective to analyze the group's adaptation process over time to the three interaction contexts designed for resolving the task (face-to-face with computer support, videoconference and e-mail). The study was a longitudinal one in which each group had to meet in several work sessions to resolve different types of task.

While it is true that this third research project shows some degree of continuity with regard to the theoretical approach of the previous ones, recent reviews of the relevant literature lead us to the consideration of alternative theoretical orientations related to new ways of understanding mediated communication. Thus, the new theoretical orientation, as we pointed out in the previous section, is based principally on two aspects: on the one hand, on the theory of Social Information Processing; and on the other, on the conception of the group as a socio-cognitive, open and multifunctional system (Walther, 1994; 1996). Our research team has incorporated this change into its most recent work (Gosálvez, in preparation; Solanes, 1999; Orengo, Zornoza, Prieto and Peiró, 2000). The results obtained confirm the importance of other variables, in addition to the technology, in the analysis of group processes and results. These other variables would include assertiveness and the social context in the prediction of uninhibited behaviour (Orengo et al., 2000), the interaction between the perceived characteristics of the technology and the objective or technical ones with regard to predicting group interaction processes (Solanes, 1999), and the importance of considering the interaction between the technology and time in the analysis of group results (Ripoll, Subirats, Torres and Marzo, 1998; Gosálvez, 1999).

It is precisely in this context, and focusing on group interaction processes, in which the objective orienting the present work emerges. Thus, based on the above considerations and the findings described, the general aim of this study is to analyze how group interaction processes change over time as a function of the technology used in the group interaction and the type of task carried out. For this purpose we formulated the following hypotheses:

Hypothesis 1. The processes related to the group's production function (coordination, involvement with the goal, motivation for effort) and those related to group well-being (positive and negative socioemotional behaviour) will present significant differences according to the communication medium. Thus, and in accordance with the literature based on theories of medium-richness, it will be the channels of least richness of information that present the lowest levels in these processes (e-mail and videoconference), except for the case of negative socioemotional behaviour, which will be higher.

However, these results will be modulated as a function of the type of task carried out by the group, in accordance with McGrath and Hollingshead's (1994) model of task-medium fitness. Thus,

- For *idea generation tasks* that require low levels of information richness, e-mail (EM) is expected to be the medium that obtains the best results in the functions of production and group well-being, followed by videoconference (VC) and face-to-face communication (FF).
- For *intellective tasks*, which require a medium level of information richness, VC is expected to be the medium that obtains the best results in the functions of production and group well-being, or it is at least expected not to present significant differences with regard to face-to-face communication.
- Finally, for *negotiation tasks* it is the groups that communicate face-to-face that are expected to develop the most positive interaction processes, given that these are complex tasks that require high levels of information richness for their successful completion. In second place for these tasks come VC and EM.

Hypothesis 2. We expect groups that use different communication media to show patterns of development over time that differ for each group function, modulated by the type of task carried out (Lebie et al. 1996; Arrow et al., 1996; McGrath and Berdhal, 1998). More specifically, we expect:

That over time groups will develop work procedures that compensate for the medium's lack of richness. Thus, the production function will improve with the passage of time in groups that communicate by e-mail, causing to disappear the differences between media in generation of ideas and intellective tasks. Meanwhile, differences may be maintained or may emerge in complex tasks that require high levels of interdependence, such as those of negotiation

That with regard to the group well-being function the differences between communication media will be maintained in the positive socioemotional behaviour perceived by the group, and will even increase in negotiation tasks. As the group develops there is more focus on the functions of well-being and member support, and less time is devoted to the production function (Lebie et al., 1996). However, we expect the opposite for negative socioemotional behaviour. In this case, with the passage of time, groups that communicate by e-mail will learn to express their disagreements or preferences by developing symbols or codes that allow them to transmit non-verbal aspects, making interaction more personal (Walther, 1996).

We shall continue with a description of the design, the sample and the variables used.

METHOD

Design, samples and procedure

In the third research project, which provides the framework for this study, we developed an experiment that involved the participation of 124 students distributed in 31 groups of 4 subjects each. Participants were Psychology students from the University of Valencia and the Jaume I University in Castellón. Groups were randomly assigned to each of the experimental conditions according to the communication channel (face-to-face with computer support, videoconference and e-mail). Also, each group worked for eight sessions according to a longitudinal experimental design. In each session they had to resolve a different task, which could be generation of ideas, intellective or negotiation. These were alternated to allow us to analyze the time effect. To this end, the sequence of the tasks was as follows: in the first two sessions and the final two, the groups performed a task of an intellective nature. In the third and fifth sessions they carried out idea generation tasks, and in the sixth session they performed negotiation tasks. At the end of each session the group members filled out a battery of questionnaires individually, and the interactions were recorded on video and/or on computer.

Variables

We shall now describe how we operationalized each of the variables included in the study: technology used for the group work, type of task, time and group interaction processes.

Communication media

The groups were distributed randomly to the different experimental conditions that made up the design, using one of the three communication media: face-to-face with computer support, videoconference and communication mediated by computer (e-mail). In *face-to-face communication* the members making up the groups were present in the same room and could use as a computer as support, to communicate with one another or to make their own notes. The sessions were filmed on video so that the information and the development of the session could subsequently be analyzed.

Interaction by means of *videoconference* was carried out in the laboratories of the Psychology Faculty in Castellón, using facilities comprising four workstations (Silicon Graphics R4600sc), equipped with a minicamera, a microphone and two speakers, and which permitted the user to connect to and disconnect from the network as he or she wished.

For the interaction by means of *e-mail* the members of each group communicated with one another through computer terminals. The software used consisted in an e-mail program installed for use with a local network. During the experiment, each member of the group was in a different section, and could only communicate with his/her colleagues via computer. The members had been introduced to one another previously, so that they knew who was in each post. They could send messages to individuals, to part of the group or to all the members at the same time.

Type of task

All the groups performed three types of tasks over the eight sessions in which they participated, following the model of Argote and McGrath (1993): generation of ideas, intellective (decision-making with correct response) and conflict of interests.

Each group had to resolve four ***intellective tasks***. In all cases these were logic problems with a single correct solution, and for which each group member had different information and the information from all of them was necessary to arrive at the correct solution. In one task they had to select the most suitable candidate to occupy the post of human resources manager in a company; in another, they had to ascertain the full names of the people who were going to be employed by a company; a third task involved correctly associating each person proposed with his or her profession, and the final task consisted in choosing among different areas of the city the most suitable location for a restaurant specializing in Valencian cuisine.

Furthermore, in another two sessions each group had to carry out a ***generation of ideas*** task. Specifically, in the third session they had to perform the task called "Slogans", based on that proposed by Pfeiffer and Jones (1980), and in the fifth session the task called "Culture Week", in which, individually, participants had to create a minimum of ten ideas for events to form part of the Psychology Faculty's Culture Week. In the second phase, participants, with their group, had to decide on a minimum of ten events that would take place.

Finally, in this experiment, all the groups performed two ***cooperation-competition*** tasks. In the fourth session, they carried out the task called "Five in a row". In this task, the four members of each group were divided into two subgroups that competed with one another in order to obtain as many points as possible. Subgroups scored highest when the participants in the game cooperated with one another. In the sixth session the members of each group were also divided into two subgroups for carrying out the task. In this case the task they had to carry out was called "Naranjas Nani". Each one of the subgroups had an interest in buying a given quantity of oranges, and had to compete with the other subgroup to obtain as many kilogrammes of oranges as possible. Once again, subgroups made most gains when they cooperated with the other subgroup.

Time

In order to analyze the extent to which an increase in experience in the use of a specific information technology by a group produces changes in group work processes, we used a longitudinal design incorporating the variable time. Specifically, we considered eight data-collection points (one for each task that each group had to resolve).

Group interaction processes

The variables considered in the present work are grouped according to McGrath's (1990) model, which highlights the multifunctionality of groups, distinguishing three functions: production (coordination of the task and involvement with the goal), well-being (negative socioemotional behaviour and positive socioemotional orientation) and personal development (motivation for effort).

- *Production function. Coordination of group members for performing the task.* This is made up of three items that indicate the extent to which the group is oriented towards carrying out the task in a structured and effective way, and how far it is coordinated and plans its efforts as a group. Its reliability is 0.76. The response scale is made up of five anchors (1-not at all, to 5-a lot)
- *Involvement in the group solution.* This is a self-report measure whose scale is composed of three items assessing the degree to which the participant feels responsible for and committed to the solution arrived at by the group. The response scale has five anchors (1-not at all, to 5-a lot). Cronbach's alpha is 0.76.
- *Motivation for effort.* This is made up of two items: "the group expected you to make an effort and become involved in the task" and "the people in the group encouraged one another mutually to achieve the greatest possible effort". Internal consistency is 0.55.
- *Group well-being function.*

Negative socioemotional behaviour.

This measure is made up of five items ("I felt frustrated or tense because of the others' behaviour", "I rejected the opinions or suggestions of the others", "I expressed negative opinions about the behaviour of others", "My opinions were rejected", and "Others expressed negative opinions about my behaviour"). Its reliability is 0.71. The response scale is made up of five anchors (1-strongly disagree, to 5-strongly agree)

Positive socioemotional behaviour.

Four items make up this variable (e.g., "The group members showed themselves to be friendly during the work session", "... everyone's opinions were heard" and "... they were respected"), with a response scale made up of five anchors (1-strongly disagree, to 5-strongly agree). Reliability is 0.76.

RESULTS

In order to tests the hypotheses proposed in the present work, we carried out several transversal and longitudinal statistical analyses, as described below.

The objective of our first hypothesis was to explore the influence of the technology used by the group on its interaction processes. In order to test it we carried out analyses of variance for each of the interaction processes considered, as a function of the communication medium used by the group, after the initial data collection. With regard to the second hypothesis, which referred to the consideration of time, understood as adaptation or experience, in the analysis of the differences in work group processes as a function of communication medium, we carried out an analysis of variance at a second point in time and a repeated-measures analysis of variance. In this way we explored possible changes in group interaction processes between the two data-collection points (first session and final session for each group) as a function of the interaction of time and communication medium. In both cases, transversal and longitudinal, the analyses are modulated by the type of task performed by the group.

The results corresponding to the different statistical tests carried out are shown in Tables 1 and 2.

First of all we shall discuss the transversal results related to the testing of Hypothesis 1. These results correspond to the data that appear in the column of Time 1 for each type of task, in Table 1.

As it can be seen in Table 1, the results obtained show statistically significant differences in some group interaction processes as a function of communication medium used. Thus, in *idea generation tasks*, members of groups that communicate face-to-face with computer support or by videoconference show higher levels of coordination among their members (F=5.14; p=0.007), more motivation for effort (F=3.89; p=0.02) and more positive socioemotional behaviour (F=7.75; p=0.001) than those of groups that communicate by e-mail. Finally, and although the required levels of significance are not reached, it is observed that participants who communicate by e-mail tend to show higher levels of negative socioemotional behaviour than those using the other two media, followed by face-to-face communication with computer support and, further behind, videoconference (F=2.91; p=0.06). These results are in the direction expected in the first hypothesis, which supports theories based on "medium-richness". However, they are not as expected according to McGrath and

Hollingshead’s (1994) task-medium fitness model. From this model we would expect that for creative tasks, which require low levels of coordination between group members, a less rich medium such as e-mail would favour group work.

As far as *intellective tasks* are concerned, the results obtained support our first hypothesis, since they indicate that the technology influences work group processes in a differential and significant way, videoconference being the medium that obtains the highest levels in coordination (F=7.15; p=0.001), involvement with the goal (F=6.77; p=0.002) and positive socioemotional behaviour (F=4.65; p=0.01), followed by face-to-face communication and, finally, e-mail. Nevertheless, it is the groups that communicate using this medium, e-mail, that perceive the highest level of negative socioemotional behaviour (F=5.60; p=0.005). These results support McGrath and Hollingshead’s (1994) task-medium fitness model.

Table 1
Analyses of variance of group interaction processes in different types of task (generation of ideas, intellective and negotiation) as a function of the communication channel use in Time 1 and Time 2.

		GENERATION OF IDEAS				**INTELLECTIVE**				**NEGOTIATION**			
		TIME 1		**TIME 2**		**TIME 1**		**TIME 2**		**TIME 1**		**TIME 2**	
		Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Task coordination	Face-to-face	4.22	0.59	4.24	0.48	4.30	0.51	3.94	0.85	3.64	0.78	4.03	0.69
	Videoconference	4.33	0.52	4.21	0.44	4.42	0.42	4.12	0.51	3.94	0.63	4.11	4.49
	E-mail	3.92	0.67	3.98	0.91	4.01	0.61	3.68	0.91	3.75	0.98	3.74	0.92
Goal involvement	Face-to-face	4.33	0.52	4.35	0.51	4.51	0.32	4.25	0.63	4.10	0.60	4.45	0.46
	Videoconference	4.46	0.37	4.42	0.39	4.52	0.30	4.47	0.46	4.17	0.60	4.43	0.51
	E-mail	4.35	0.56	4.17	0.77	4.23	0.54	4.17	0.65	3.97	0.70	4.23	0.66
Level of effort	Face-to-face	4.08	0.56	3.88	0.53	3.90	0.66	3.91	0.84	3.58	0.86	3.99	0.53
	Videoconference	4.06	0.49	3.85	0.55	4.08	0.71	3.86	0.65	3.78	0.61	3.97	0.61
	E-mail	3.76	0.67	3.91	0.84	4.00	0.56	3.81	0.66	3.77	0.78	3.75	0.64
Neg. socioemotional behvr.	Face-to-face	1.74	0.59	1.75	0.67	1.66	0.58	1.64	0.74	1.94	0.86	1.78	0.68
	Videoconference	1.72	0.64	1.60	0.59	1.68	0.52	1.50	0.61	1.70	0.81	1.54	0.53
	E-mail	1.94	0.73	1.90	0.67	2.07	0.66	1.82	0.63	2.04	0.78	2.16	0.82
Pos. socioemotional	Face-to-face	4.76	0.32	4.56	4.46	4.58	0.55	4.38	0.66	4.33	0.55	4.33	0.57
	Videoconference	4.71	0.33	4.58	0.39	4.70	0.38	4.53	0.41	4.60	0.52	4.55	0.48
	behvr.	E-mail	4.44	4.49	4.31	0.67	4.39	0.47	4.24	0.59	4.06	0.70	4.01

Table 2
Longitudinal analyses of group interaction processes in different types of task (generation of ideas, intellective and negotiation)

	GENERATION OF IDEAS						**INTELLECTIVE**						**NEGOTIATION**					
	Time		**Channel**		**T * C**		**Time**		**Channel**		**T * C**		**Time**		**Channel**		**T* C**	
	F	**p**	**F**	**p**	**F**	**p**	**F**	**p**	**F**	**p**	**F**	**p**	**F**	**p**	**F**	**p**	**F**	**p**
Task coordination	0.07	0.79	**4.48**	**0.001**	0.86	0.43	**25.9**	**0.001**	**6.50**	**0.002**	0.06	0.94	**5.27**	**0.02**	2.33	0.10	1.84	0.16
Goal involvement	1.99	0.16	1.6	0.21	1.52	0.22	**5.31**	**0.02**	**5.57**	**0.005**	1.63	0.20	**26.55**	**0.001**	1.72	0.18	0.25	0.78
Level of effort	2.64	0.11	0.81	0.45	**4.81**	**0.01**	3.39	0.07	0.18	0.84	1.00	0.37	**6.91**	**0.01**	0.54	0.58	**3.02**	**0.05**
Neg. socioemotional behvr.	**10.71**	**0.001**	**7.07**	**0.01**	0.24	0.79	**10.91**	**0.001**	**5.41**	**0.006**	0.10	0.91	0.26	0.61	**13.59**	**0.001**	0.07	0.94
Pos. socioemotional behvr.	0.40	0.50	**2.91**	**0.06**	0.34	0.71	**5.90**	**0.02**	**5.60**	**0.005**	1.18	0.36	0.52	0.47	**7.28**	**0.001**	1.44	0.24

T * C = Interaction Time X Channel

Finally, and in the case of *negotiation tasks*, statistically significant differences are only found in positive socioemotional behaviour as a function of the technology used by the group. Thus, participants in the videoconference condition perceive the highest levels of positive socioemotional behaviour, followed by face-to-face communication with computer support and, further behind, e-mail (F= 8.58; p=0.001). These results indicate that for tasks requiring high levels of coordination between group members and the interchange of different ideas and interests, the communication medium used by the group does not have a significant influence on the way the group interacts. These findings do not support our first hypothesis.

We shall continue by discussing the results obtained in the analyses carried out to test Hypothesis 2. According to this hypothesis, we expected the communication media to show different patterns of development over time for each group function modulated by type of task. The results we shall discuss appear in Table 1, in the column Time 2, and in Table 2. These data reflect the changes experienced by the group with the passage of time (comparing the scores of Time 1 with those of Time 2) and in interaction with the communication medium used.

With regard to the *production function* we expected that with the passage of time the groups would develop procedures that allowed them to compensate for the lack of richness of the medium, being effective in performing the task in a medium with less richness than necessary. That is, we expected the differences due to the technology to disappear for creative tasks and intellective tasks, and for them to persist or appear in tasks with a high level of interdependence, such as those of negotiation.

As can be seen in Table 1, for *idea generation tasks* the results indicate that the differences obtained as a function of the communication medium disappear on comparing the first time point with the second.

The longitudinal analyses, shown in Table 2, indicate that if we consider the scores of Time 1 and Time 2 jointly, there are significant differences according to the communication channel in coordination (F=4.48; P=0.001). The interaction between time and the communication medium is only statistically significant for motivation for effort (F=4.81; p=0.01). This result indicates that group members' perceptions with regard to motivation for effort evolve over time in a different way depending on the communication channel used. Thus, the groups that communicate by e-mail are those that feel more motivated to make an effort as time passes. However, the groups that use videoconference or communicate face-to-face with computer support perceive a lower level of motivation for effort as their experience increases (see Figure 1).

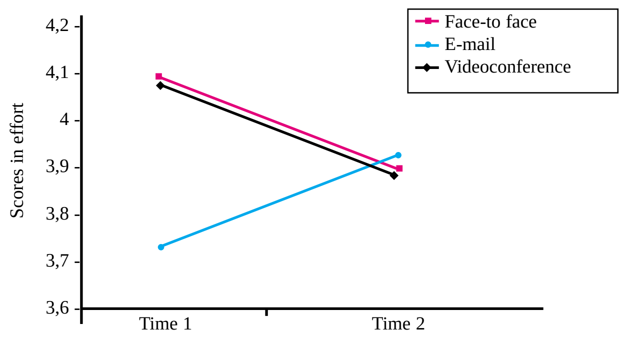
Thus, only motivation for effort increases with experience in the use of e-mail. Neither coordination level nor involvement with the goal increase, though the latter does not present significant differences as a function of the communication medium at either of the two time points considered.

As far as *intellective tasks* are concerned, the differences in the production function between the communication media are maintained with the passage of time, that is, if we compare Time 1 and Time 2 (Table 1), so that the expectations of the second hypothesis are not fulfilled. Thus, it continues to be the rich media, videoconference and face-to-face, that obtain the highest scores, at the second time point, in coordination (F=3.40; p=0.04) and in group members' level of involvement with the goal set (F=3.07; p=0.05).

If we consider the scores of Time 1 and Time 2 jointly (longitudinal analyses, Table 2), there appear significant differences in the production function as a function of both time and channel. In general, with the passage of time scores in all the group interaction processes decrease, perhaps because the groups become accustomed to performing these types of task. Even so, in none of the processes studied did we find a significant interaction between time and the communication medium used in this type of task, so that our second hypothesis was not confirmed.

As regards the *negotiation tasks*, the results obtained

Figure 1
Evolution of group effort level as a function of communication channel in generation of ideas tasks



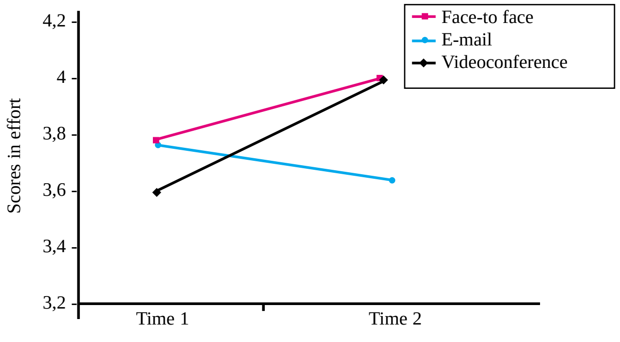
support Hypothesis 2, since in this case, at the first time point there are no significant differences in the production function according to the technology used (Hypothesis 1). However, at the second time point such differences do appear, with groups coordinating better in videoconference and face-to-face than in e-mail (F=3.08; p=0.05).

The longitudinal analyses show a significant effect of time on the coordination of group members (F=5.27; p=0.02), on involvement with the goal (F=26.55; p=0.001) and on motivation for effort (F=6.91; p=0.01). Likewise, we found a statistically significant time-technology interaction effect for motivation for effort, as occurred in the intellective tasks. This variable presents different trajectories over time in the groups that use different communication media. Thus, those participants who communicate face-to-face with computer support or by means of videoconference perceive themselves as more motivated to make an effort as their experience increases, whilst those that use e-mail practically maintain their perception with the passage of time and with increased experience (see Figure 2). It can be said, therefore, that the communication media present different trajectories of development and the differences with respect to e-mail increase, a finding that supports Hypothesis 2.

With regard to the *group well-being* function, we expected the differences between communication media to persist in the positive socioemotional behaviour perceived by the group, and that they would even increase in negotiation tasks; for negative socioemotional behaviour we expected the opposite.

In idea generation tasks at the second time point, the differences in positive socioemotional behaviour perceived by the group continue to be statistically significant (F=3.66; p=0.03). In this case, they follow the direction expected by the "social cues filtering" theories, so that the groups communicating by e-mail continue to present the lowest level, in contrast to the results expected by Walther (1996). The longitudinal analyses, shown in Table 2, indicate that if we consider the scores of Time 1 and Time 2 jointly, there are significant differences according to communication channel in the well-being function: positive socioemotional behaviour (F=7.07; p=0.001) and near-significance in negative socioemotional behaviour (F=2.91; p=0.06). There are only differences as a function of time in positive socioemotional behaviour (F=10.71; p=0.001). The interaction between time and communication medium is not statistically significant. Thus, Hypothesis 2 is confirmed as far as this type of task is concerned, with regard to perceived positive socioemotional behaviour.

Figure 2
Evolution of group effort level as a function of communication channel in negotiation tasks



As far as *intellective tasks* are concerned, the differences are maintained in positive socioemotional behaviour as a function of the communication medium over time (F=2.98; p=0.05), that is, if we compare Time 1 and Time 2 (Table 1), so that the expectations of the second hypothesis are fulfilled. It continues to be the rich media, videoconference and face-to-face, that obtain the highest scores. Moreover, the significant differences in negative socioemotional behaviour disappear, since they do not attain the conventional significance level (F=2.45; p=0.09). Thus, when the groups become accustomed to performing this type of task via e-mail they change their perception of the socioemotional behaviour expressed by the group members, perceiving it less negatively.

When we consider the scores of Time 1 and Time 2 jointly (longitudinal analyses, Table 2), we find significant differences in the well-being function according to both time and channel. Even so, in none of the processes studied do we obtain a significant interaction between time and the communication medium used in this type of task, so that our second hypothesis is confirmed.

With regard to *negotiation tasks*, the results obtained also support Hypothesis 2. The differences found as a function of the communication medium for positive socioemotional behaviour at the first time point are maintained at the second point (F=7.98; p=0.001). Furthermore, there appear significant differences for the perception of negative socioemotional behaviour (F=8.66; p=0.001), with e-mail being the channel that presents the highest levels of this type of behaviour, so that these results are not in the direction expected in our second hypothesis (Walther, 1996).

The longitudinal analyses show a significant effect of the communication channel on the group's socioemotional behaviour (positive socioemotional behaviour (F=13.59; p=0.001) and negative (F=7.28; p=0.001)). We did not find significant differences on considering the interaction time X communication channel. Subsequently, Hypothesis 2 is confirmed for positive socioemotional behaviour, and the opposite occurs for negative socioemotional behaviour, since the differences between media emerge over time.

DISCUSSION

As the reader will recall, the objective of the present work consisted in analyzing how group interaction processes change across time as a function of the technology used in the group interaction and the type of task carried out. This objective was divided into two hypotheses.

The first was based on theories of the richness of the communication media according to the type of information they allowed group members to transmit (Short et al., 1976; Daft and Lengel, 1986). According to these theories, the communication medium influences group work in such a way that rich channels (face-to-face or videoconference) favour it and poorer channels, such as e-mail, make it more difficult, as they filter non-verbal and social information (Kiesler et al., 1984; Sproull and Kiesler, 1992; Siegel et al., 1986). Nevertheless, McGrath and Hollingshead (1994) point out that the influence of the technology on group work will depend on the type of task to be performed, and more specifically, on the degree of fit between the richness offered by the medium and that demanded by the task.

Our results support McGrath and Hollingshead's (1994) idea that the influence of the technology on a group's work depends on the type of task it has to perform, but they are not always in the direction hypothesized by these authors. In idea generation tasks the results obtained show that groups coordinate themselves worse, express less motivation for effort and develop less positive socioemotional behaviour when they communicate by e-mail than when they do so face-to-face or via videoconference, even though the task may have low demand for richness of information. But the task-medium fitness model is indeed confirmed in intellective tasks. These types of task require a medium level of richness of information, and it is videoconference that presents the highest scores in the functions of production and group well-being. On the other hand, the results obtained for negotiation tasks do not support this model. In this type of task the communication medium only has a significant influence on the positive socioemotional behaviour perceived by the group, which is higher for videoconference and lower for e-mail.

In general, it can be said that the results obtained are in the direction proposed by the medium-richness theories. The richer media favour the functions of production and group well-being, whilst the poorer media, such as e-mail, continue to present lower levels. However, these effects are modulated by the type of task the group is performing, so that the influence of the technology differs across the three types of task considered in our study. When groups have to perform a task of an intellective type they work better via videoconference, since the level of richness offered by this medium coincides with that required by the task.

Our second hypothesis proposed that communication media present different patterns of development over time for each group function, modulated by type of task. Patterns of change are neither simple nor consistent. The differences between media may persist over time, may decrease or emerge, or different patterns of development may be found (Arrow et al., 1996).

With regard to the *production function*, we expected that with the passage of time groups would develop procedures that allowed them to compensate for the lack of richness of the medium, performing the task effectively in a medium with less richness than necessary (McGrath and Berdhal, 1998). Thus, we expected differences dependent upon the technology to disappear for creative and intellective tasks, whilst differences would be maintained or appear in tasks involving high levels of interdependence, such as those of negotiation.

This hypothesis is confirmed in creative and negotiation tasks, but not in those of an intellective nature. With regard to idea generation tasks, the results show that when groups acquire experience in the use of communication media, differences in the production function disappear. With regard to motivation for effort, this presents different trajectories according to time and channel. Groups that communicate via e-mail increase their motivation for task-related effort with the passage of time, whilst such motivation decreases in the richer media.

However, when groups perform tasks that require the sharing and integration of the information possessed by each member (intellective tasks), group work continues to be more positive in the rich media than in e-mail across time. That is, the differences between media persist over time.

Negotiation tasks present different results from the above. With the passage of time there emerge differences according to the communication medium used, so that groups coordinate themselves better in rich media than when they use e-mail. Furthermore, motivation for effort presents different trajectories as a function of time and communication medium, so that groups that communicate by means of media that allow for greater richness of information (face-to-face and videoconference) increase their motivation as they acquire experience, whilst for those that use e-mail their perception remains constant, in contrast to what occurred in the creative tasks. These results may indicate that groups need more time to adapt to the technology when they perform more complicated tasks, as they need to integrate different ideas and points of view. It may be that they first need to become accustomed and develop strategies for completing the task, so that differences do not emerge according to the communication medium; rather, it is later, when the group has developed strategies for achieving the objective, that the influence of the technology becomes significant. The main effect of time is significant in the production function, whilst the main effect of channel is significant in the well-being function.

With regard to the *group well-being* function, we expected the differences between communication media to be maintained in the socioemotional behaviour perceived by the group, and even to increase in negotiation tasks. Lebie et al. (1996) point out that the patterns of change in socioemotional behaviours in face-to-face communication and in mediated communication suggest that several causal factors may be involved. The initial differences may be due to the novelty of the medium and the slowness of writing messages in e-mail communication. But with the passage of time, socioemotional behaviour increases in both media, so that more factors are involved. These authors point out that there are main effects of the technology and time, but not of the interaction between the two. This indicates that groups change their interaction patterns in a similar way in the two media, and that there may be differences that persist over time.

On the other hand, we expected the contrary to be the case with negative socioemotional behaviour, bearing in mind the results reviewed by Walther (1996) and the theory of social information processing. Walther stresses the importance of experience in the use of mediated communication so that users perceive it in a more positive way and so that they learn how to develop positive socioemotional behaviour in its context.

Our results for generation of ideas and intellective tasks support the hypothesis, since the differences are maintained in positive socioemotional behaviour (the rich media being those that present the highest scores) and disappear in negative socioemotional behaviour; in negotiation tasks, significant differences are maintained in both types of socioemotional behaviour, positive and negative. That is, in this more complex type of task, the differences between media in negative socioemotional behaviour do not decrease solely with experience in use of the medium.

In general, our results show that although the technology used has a significant influence on group work, it is not only its objective characteristics that change work group processes; rather, it is the interaction of the technology with other factors, such as the task or experience, that affect the differences that appear in the interaction processes of a group.

The type of task performed by the group appears as an important modulating factor in the relationship between the technology and group work. It is important to bear in mind that a fit must be found between the characteristics of the task and those offered by the technology, though our results only support McGrath and Hollingshead's (1994) task-medium fitness model for intellective tasks. However, this fit is not static but dynamic, changing over time, so that it is the result of a triple interaction: task, technology and experience (time) (McGrath and Berdhal, 1998). The result of this interaction is complex, and there is a need to operationalize and measure more accurately the lack of fit found in the model. In our study, negotiation (competitive) tasks show the initial importance of time for establishing habitual procedures of group work, in order to subsequently analyze the differences introduced by the technology and seek a situation of fit.

It is also important to highlight the influence of the technology on the well-being function of the group, in this case on the socioemotional aspects of group interaction. For the three types of task, e-mail is perceived as the medium in which least positive behaviours emerge in comparison to face-to-face communication and videoconference, and these differences are maintained with the passage of time. This result supports the filtering of cues theories, and does not confirm those of Strauss

(1997) or Walther (1995; 1996). That is, the continued use of e-mail does not make the group perceive it more positively (Walther, 1996), even though it does reduce negative socioemotional behaviour. It may be that the group becomes accustomed to using the technology and learns to express its disagreements with or its rejections of others' ideas in a less explicit or less negative way, while it needs more time to develop new strategies that permit it to better express and transmit positive socioemotional aspects.

Thus, our results are in support of McGrath and Berdhal's (1998) call for a differentiation between the effects of the technology itself and those of the "novelty" represented by the use of that technology. Secondly, our study confirms the importance of analyzing the specific effects on work processes and results of the objective characteristics of the technology, effects that may persist and even increase over time, as well as those resulting from the interaction between the technology, the task, time and other factors, such as the group's perception of the communication medium it is using (Solanes, 1999). And thirdly, the present work highlights the need to take care in the design of the measures for the different functions and results of a group. In this regard, each technology offers different opportunities and constraints for the interaction of group members. Thus, the measure of coordination level, of conflict management or of the way of expressing agreements or disagreements does not have to be the same for all the communication media (Zornoza, 1992). Future studies might include measures of observation of group functioning combined with self-report measures, which would permit a more detailed analysis of the strategies developed by groups and of their evolution as a function of time and as a function of the technology used.

ACKNOWLEDGEMENTS

The present study was subsidized by the research grant: *Trabajo cooperativo en grupos mediados por nuevas tecnologías. Un estudio longitudinal*, from the Generalitat Valenciana (GV-3239/95).

REFERENCES

Acín, C. (1995). *Aspectos psicosociales de la toma de decisiones en grupo que utilizan comunicación mediada por ordenador y videoconferencia*. Doctoral thesis. Universidad de Valencia.

Alcover, C.M. & Gil, F. (in press). Influencias de las nuevas formas de flexibilidad organizacional sobre los equipos de trabajo. *Revista de Psicología General y Aplicada.*

Andriessen, E. & Van der Velden, J. (1993). Team work supported by interaction technology. Communication presented at 6th European Congress on Work and Organizational Psychology. Alicante, April, 1993

Argote, L. & McGrath, J. (1993). Group processes in organizations: continuity and change. In C.L. Cooper & I.T. Robertson (Eds.): *International Review of Industrial and Organizational Psychology, 8.* John Wiley & Sons, New York.

Arrow, H.; Berdahl, J; Bouas, K; Craig, K; Cummings, A.; Lebie, L.; MacGrath, J.; O'Connor, K.; Rhoades, J. & Schlosser, A. (1996). Time, Technology and groups: An integration. *Computer Supported Cooperative Work, 4*, 251-261.

Bales, R.F. (1950). *Interaction process analysis: A method for the study of small groups*. Reading, MA. Addison Wesley.

Daft, R.L. & Lengel, R.H. (1986). Organizational information requirement, media richness and structural design. *Management Science*, 32, *5*, 554-571.

Daft, R.L.; Lengel, R.H. & Trevino, L.K. (1987). Message equivocality, media selection and manager performance: Implication for information systems. *MIS Quarterly, 11*, 354-366.

DeSanctis, G. & Poole, M.C. (1997). Transitions in teamwork in new organizational forms. *Advances in Group Processes, Vol. 14*, pp. 157-176.

Dubrovsky, V. J.; Kiesler, S. & Sethna, B. N. (1991). The equalization phenomenon: Status effects in computer-mediated and face-to-face decision-making groups. *Human Computer Interaction, 6,* 119-146.

Gosálvez, I. (1999). *Incidencia del medio de comunicación sobre la eficacia grupal. Un estudio longitudinal*. Degree dissertation. Universidad de Valencia.

Guzzo, R.A.; Yost, P.R.; Campbell, R.J. & Shea, G.P. (1993). Potency in groups: articulating a construct. *British Journal of Social Psychology, 32*, 87-106

Hackman, J.R. (1987). The design of work teams. In J.W. Lorsch (Ed.). *Handbook of organizational behavior*, 315-342. Prentice Hall. Englewood Cliffs.

Hackman, J.R. & Morris, C.G. (1975). Group tasks, group interaction processs and group performance effectiveness: A review and proposed integration. In L. Berdowitz (Ed.) *Advances in Experimental Social Psychology, 8,* 45-99. Academic Press. New York.

Hiltz, S.R.; Johnson, K. & Turoff, M. (1986). Experiments in Group Decision Making: Communication Process and Outcome in face-to-face versus Computerized Conferences. *Human Communication Research, 13*, 225-252.

Hollingshead, A.B.; McGrath, J.E. & O'Connor, K.M. (1993). Group Task Performance and Communication Technology. A Longitudinal Study of Computer-Mediated versus Face-to Face Work Groups. *Small Group Research*, 24, *3*, 307-333.

Hollingshead, A.B. & McGrath, J.E. (1995). Computer assisted groups: A critical review of the empirical research. In Guzzo, R.; Salas, E. (De.). *Team effectiveness and decision making in organization*s. Jossey Bass Publishers. San Francisco.

Huber, G.P. (1990). A theory of the effects of advanced information technologies on organizational design. *Academy of Management Review, 15*, 47-71.

Johansen, R. (1988). *Groupware: Computer support for business teams.* New York: Free Press.

Kiesler, S. & Sproull, L. (1992). Group decision making and communication technology. *Organizational behavior and human decision processes, 52*, 96-123.

Kiesler, S.; Siegen, J. & McGuire, T. (1984). Social Psychological Aspects of Computer Mediated Communication. *American Psychologist*, *39*, 1123-1134.

Lebie, L.; Jonathan, R. & McGrath, J. (1996). Interaction Process in Computer mediated and Face to Face groups. *Computer Supported Cooperative Work, 4*, 127-152.

Marzo, J.C.; Rodríguez-Molina, I.; Ripoll, P. & Marín, F. (1995). La influencia del canal de comunicacion y el tipo de tarea sobre el funcionamiento socio-emocional del grupo. *Communication presented at V Congreso Nacional de Psicologia Social*, Salamanca, 18-20 December.

McGrath, J. (1984). *Group interaction and performance*. Englewood Cliffs. NJ, Prentice Hall.

McGrath, J. & Hollingshead, A. (1994). *Groups interacting with technology*. Sage Publications.

McGrath, J.E (1990). Time matters in groups. In J. Galegher, R.E. Kraut, & C. Egido (Ed.). *Intellectual teamwork: Social and technological foundations of cooperative work*. Hillsdale, NJ: Lawrence Erlbaum.

McGrath, J.E (1991). Time, interaction, and performance (TIP): A theory of groups. *Small group research*, *22*, 147-174.

McGrath, J.E. & Arrow, J. (1996). Introduction: The JEMCO-2 Study of Time Technology, and Groups. *Computer Supported Cooperative Work* (CSCV), *4*, 107-126.

McGrath, J.E. (1993). Introduction: The JEMCO Workshop. Description of a longitudinal study. *Small Group Research*, *24*, 285-306.

McGrath, J.E. & Berdhal, J.L. (1998). Groups, Technology, and Time. Use of computers for collaborative work. In R. Scott Tindale et al. (Eds.), *Theory and Research on Small Groups*. Plenum Press, New York.

McGrath, J.E. & O'Connor, K.M. (1996) Temporal issues in work groups. In M. West. (Ed.). *Handbook of workgroup psychology,* (pp. 25-52). London, Wiley.

Mennecke, B.E.; Hoffer J.A. & Wynne, B.E. (1992). The implications of group development and history for group support system theory and practice. *Small Group Research*, Vol. 23, *4*, 524-572.

Nelson, J.L. (1995). *Post-Industrial Capitalism.* Beverly Hills. Sage.

O'Connor, K; Gruenfeld, D. & McGrath, J. E. (1993). The experience and effects of conflict in continuing work groups. *Small Group Research*, 24, *3*, 362-382.

Orengo, V. (1994). *Implicaciones de la conducta desinhibida en contextos de comunicación grupal mediada por nuevas tecnologías*. Degree dissertation, Universitat de València.

Orengo, V.; Gosálvez, Y.; Fernández, M.C. & Prieto, F. (1995). Análisis de la interacción grupal en diferentes canales de comunicación a través de la técnica IPA. In J.C. Sánchez & A.M. Ullán (Comps.): *Procesos Psicosociales Básicos y Grupales*, 501-515. Ed. Eudema, Salamanca.

Orengo, V.; Zornoza, A.; Acín, C.; Prieto, F. & Peiró, J.M. (1996). Análisis de la interacción grupal a través de medidas de observación en comunicación mediada. *Revista de Psicología Social*, 11, *2*, 151-162.

Orengo, V.; Zornoza, A.; Prieto, F. & Peiró, J.M. (2000). The influence of familiarity among group members, group atmosphere and assertiveness on uninhibited behavior through three different communication media. *Computers in Human Behavior*, *16*, 141-159.

Peiró, J.M.; Prieto, F. & Zornoza, A. (1993). Nuevas tecnologías telemáticas & trabajo grupal. *Psicothema*, *5*, 287-305.

Peiró, J.M.; Prieto, F. & Zornoza, A. (1994). The role played by interaction technology perceived characteristics in group mediated communication. Work presented at 23rd International Congress of Applied

Psychology, Madrid.

Peiró, J.M.; Prieto, F.; Zornoza, A. & Ripoll, P. (1999). El conflicto y su gestión en el trabajo en equipo mediado por nuevas tecnologías de la información. In Morales, F. & Yubero, S. (Coords.), *El grupo y sus conflictos*. Edit. Universidad de Castilla-La Mancha, Cuenca.

Pfeiffer & Jones (1980). *User's guide to structured experience kit*. University Associates Inc. Santiago, California.

Poole, M.S.(1991). Procedures for managing meetings: Social and technological innovation. In R. Swanson & B. Knapp (Eds.) *Innovative meeting management*, pp. 53-110. Austin, TX: 3M Meeting Management Institute.

Short, J.; Williams, E. & Christie, B. (1976). *The Social Psychology of Telecommunications*. Edit. John Wiley & Sons. London.

Siegel, J.; Dubrosvsky, V.; Kiesler, S. & McGuire, T.W. (1986). Group Processes in Computer mediated Communication. *Organizational Behavior and Human Decision Processes*, Vol. *37*, pp. 157-198.

Siegel, J.; Dubrovsky, V.; Kiesler, S. & McGuire, T.W. (1986) Group Processes and Computer-Mediated Communication. *Organizational Behavior and Human Decision Processes*, *37*, 157-187.

Smolensky, M.W.; Carmody, M.A. & Halcomb, C.G. (1990). The Influence of Task Type, Group Structure and Extraversion on Uninhibited Speech in Computer-Mediated Communication. *Computers in Human Behavior*, *6*, 261-272.

Solanes, A. (1999). *Nuevas tecnologías y trabajo en grupo: la incidencia de la percepción del medio de comunicación*. Doctoral thesis. Universidad de Valencia.

Spears, R. & Lea, M. (1992). Social influence and the influence of the social in computer-mediated communication. In M. Lea, Ed., *Contexts of computer-mediated-communication.(2-65)*. Hemel Hempstead: Harvester-Wheatsheaf.

Steiner, I.D. (1972). *Group process and productivity*. New York, Academic Press.

Straus, S.G. (1997). Technology group process and group outcomes: testing the connections in computer mediated and face to face groups. *Human Computers Interaction*, *12*, 227-266.

Straus, S.G. & McGrath, J.E. (1994) Does the medium matter: The interaction of task and technology on group performance and member reactions. *Journal of Applied Psychology*, *79*, 87-97.

Sundstrom, E.; De Meuse, K.P. & Futrell, D. (1990). Word teams: Applications and effectiveness. *American Psychologists*, *45*, 120-123.

Torres, M.J; Peiró, J.M; Acín, C & Zornoza, A.(1995). Análisis del funcionamiento grupal orientado a la tarea en función de la tecnología y el tipo de tarea empleados. *Communication presented at V Congreso Nacional de Psicologia Social*, Salamanca, 18-20 December.

Walther , J.B. (1992). A longitudinal experiment on relational tone in computer-mediated and face to face interaction. In J.F. Nunamaker Jr. & R.H. Sprague (Eds.), *Proceedings of the Twenty-Fifth Annual Hawaii International Conference on System Science*, *4*, 220-231. Los Alamitos, CA:IEEE Computer Scoiety Press.

Walther, J.B. (1994). Anticipated ongoing interaction versus channel effects on relational communication in computer-mediated interaction. *Human Communication Research*, *40*, 473-501.

Walther, J.B. (1995). Relational aspects of computer-mediated communication: Experimental observations over time. *Organization Science*, *6*, 186-203.

Walther, J.B. (1996). Computer-mediated communication: Impersonal, interpersonal and hyperpersonal interaction. *Communication Research*, *23*, 3-43.

Walther, J.B. & Burgoon, J.K. (1992). Relational communication in computer-mediated interaction. *Human Communication Research*, *19*, 50-88.

Zornoza, A.; Acín, C.; Marín, C.; Hontangas, P. & Prieto, F. (1994). Incidencia de las características percibidas del canal de comunicación sobre distintos aspectos del proceso grupal. In Clemente, A.; Musitu, G. & Gutierrez, M. (Dir.) *Intervención educativa y desarrollo humano*. Edit. Cristóbal Serrano, Valencia.

Zornoza, A.; Prieto, F.; Martí, C. & Peiró, J.M. (1995). Group productivity and telematic communication. *European Work and Organizational Psychologist*, 3, *2*, 117-127.

Zornoza, A. (1992). *La toma de decisiones en grupo. Efectos diferenciales del canal de comunicación sobre el rendimiento grupal*. Doctoral thesis, October. Valencia.

Zornoza, A.; Ripoll, P. & González, P. (1995). La percepción del canal de comunicación: Un estudio de sus antecedentes personales y de su incidencia sobre el resultado grupal. In González, L; De la Torre, A. & De Elena, J. (Comps.), *Psicología del Trabajo y de las Organizaciones, Gestión de Recursos Humanos y Nuevas Tecnologías*. Edit. Eudema, Salamanca.

Psychology in Spain, 2002, Vol. 6. No 1, 56-60

FAMILY INTERVENTION PROGRAMME IN SCHIZOPHRENIA: TWO-YEAR FOLLOW-UP OF THE ANDALUSIA STUDY

José A. Muela Martínez*and Juan F. Godoy García**
*University of Jaén **University of Granada

Se presentan los resultados tras dos años de seguimiento del estudio de Andalucía, un programa de intervención familiar en esquizofrenia basado en la reducción de la Emoción Expresada, la disminución del Estrés Familiar y el aumento del nivel de conocimientos sobre la esquizofrenia por parte de los familiares que ya mostró su eficacia tras un año de seguimiento. Participan 25 familias (una muerte experimental del 4% sobre el primer año de seguimiento). La tasa de recaídas en el grupo experimental es del 0% frente al 40% del grupo control. Se comprueba, así, que la eficacia del programa se mantiene durante dos años. También se comparan los datos de este estudio con los de otros programas de intervención familiar en esquizofrenia que informan de seguimiento de dos años.

Following two years of monitoring of the Andalusia Study, this article presents the results of a programme of family intervention in schizophrenia based on reduction of Expressed Emotion, reduction of Family Stress and increase in relatives' knowledge about schizophrenia. The programme had already shown its effectiveness on a one-year follow-up. Twenty-five families participated in the present study (an attrition rate of 4% with respect to the first year of monitoring). The relapse rate in the experimental group was 0% versus 40% in the control group. The effectiveness of the programme over the two-year period can, therefore, be sustained. The data from this study are compared with those of other family intervention programmes in schizophrenia that report on monitoring over two years.

The 1980s saw the emergence of a series of studies which, using family intervention in schizophrenia, attempted – mostly successfully – to reduce the relapse rate in such patients. All of these studies were based on Zubin and Spring's (1977) stress-vulnerability theory, and the majority on the Expressed Emotion construct (Brown, Birley and Wing, 1972; Vaughn and Leff, 1976). Expressed Emotion (EE) is a form of interaction between a schizophrenic's relative and the schizophrenic him/herself, involving, on the part of the former, some or all of the following characteristics: criticism of the patient's behaviour (in content or tone of voice), generalized hostility towards or rejection of him/her as a person, and emotional over-involvement (overprotection, self-sacrifice, desperation or intense emotional reactions). When one family member presents EE, the whole family is considered as being high-EE. Today it is admitted that a schizophrenic who lives with a high-EE family has four times more probability of relapse than a patient whose family is low-EE.

According to the stress-vulnerability theory, EE is a stressor capable of producing relapse in a vulnerable subject (and a recovered schizophrenic is such a subject). The objective of the family intervention is that of reducing EE as a way of controlling the level of stress the subject suffers and thus avoiding relapse. Studies that have successfully used family intervention in schizophrenia are:

The Ventura Study (Goldstein, Rodnick, Evans, May and Steinberg, 1978): This study developed a crisis-oriented therapy, informing the family about the illness and its relationship with stress and teaching it to detect possible stressors and how deal with them. Its methodology is similar to that of Problem-Solving, but it is unstructured. It does not take into account EE. Its results six months after patients' discharge were: 0% of relapses in the experimental group versus 48% in the control group.

The Camberwell Study (Leff, Kuipers, Berkowitz, Eberlein-Vries and Sturgeon, 1982; Leff, Kuipers, Berkowitz, Eberlein-Vries and Sturgeon, 1983; Leff, Kuipers, Berkowitz, Eberlein-Vries and Sturgeon 1985): After four in-home information sessions with the families, they were invited to sessions of groups of families (without the presence of the patient) in order to talk about the problems of living with a schizophrenic and how to deal with them. Strategies used were an increase in the family's social support network and a decrease in feelings of family isolation. Sessions were also organized with individual families (with the patient present),

---

The original Spanish version of this paper has been previously published in *Apuntes de Psicología*, 2001, No 3, 421-430

....

Correspondence concerning this article should be addressed to José A. Muela, Departamento de Psychología, Facultad de Humanidades y CC de la Education (Edf. D-2). Campus "Las Lagunillas". 23071 Jaén. Spain. E-mail: jmuela@ujaen.es

though with a highly unstructured format. The results of this study nine months after patients' discharge were: 8% of relapses in the experimental group versus 50% in the control group.

The California Study (Falloon, Boyd and McGill, 1984, Falloon, Boyd, McGill, Razani, Moss and Gilderman, 1982; Falloon, Boyd, McGill, Strang and Moss, 1981): After a few educational sessions for the whole family the programme began. A structured problem-solving method was used, in the family home with each family member individually and with the patient present. The results after patients' discharge were: 6% of relapses in the experimental group versus 44% in the control group.

The Pittsburgh Study (Anderson, Reiss and Hogarty, 1986; Hogarty, Anderson, Reiss, Kornblith, Greenwald, Javna and Madonia, 1986): After a short but intensive period in which families are informed about the illness, the family is provided with a series of techniques to improve its emotional climate by means of a highly structured programme, emphasising reduction of stress, reduction of guilt feelings, widening of social networks, gradual increase in the patient's responsibilities, and so on. Results nine months after patients' discharge were: 9% of relapses in the experimental group versus 28% in the control group.

The Salford Study (Barrowclough and Tarrier, 1990; Tarrier, Barrowclough, Vaughn, Bamrah, Porceddu, Watts and Freeman, 1988): After a brief period of education about the illness, families are trained in the planning of goals and in behavioural coping techniques designed to reduce stress. Results nine months after patients' discharge were: 12% of relapses in the experimental group versus 53% in the control group.

The Birmingham Study (Mentioned in Tarrier and Birchwood, 1995): As a criterion for selecting families, this study uses not high family EE but subjective family burden. This intervention is psychoeducational, and focuses on relief of family burden and resolution of parents' feelings of loss. No data are available for relapse rates, but they are not statistically different between the groups. It would appear that the dependent variable of this study is subjective perception of burden, which was reduced in the experimental group after the programme.

The Andalusia Study (Muela and Godoy, in press). This is based on the reduction of Expressed Emotion (EE) and Family Stress, as well as on the increase of knowledge about the illness on the part of family members as a means of reducing relapse rates in schizophrenics. The techniques used were, among others, problem-solving, relaxation, modification of irrational thoughts, development of communication skills and family counselling. In the follow-up one year after the start of the programme, the relapse rate in the experimental group was 20%, compared to 63.3% in the control group.

In longer follow-ups, except that of the Ventura Study (see Goldstein and Kopeikin, 1981) and the Birmingham Study (no data), all the projects mentioned obtained significantly lower relapse rates in the experimental group than in the control group (this issue is treated in more depth in the Discussion section). In the present study we examine the effectiveness of the Andalusia Study after two years' monitoring of the original sample.

METHOD

Subjects and Groups

Participants in the research were the families that had formed part of the Andalusia Study (all parents of patients, except for one, the wife of a patient): the fifteen from the experimental group and ten of the eleven from the control group (it was impossible to locate one control group family due to a change of residence). Their characteristics are shown in Table 1.

Procedure

The Andalusia Study consists of four phases:

- Phase 0 (*Evaluation*): selection of participants in the programme. Those selected for the experimental passed to the next phase; those of the control group to Phase III (Follow-up). Assessment of the families was made by means of the Camberwell Family Interview (CFI) in its Spanish version by Gutiérrez (1986); family EE level was assessed using a questionnaire on knowledge about schizophrenia designed for this study and Álvarez and Gutiérrez's

Table 1
Characteristics of the subjects in the Andalusia Study
(taken from Muela, 1999)*

VARIABLES	EXPERIMENTAL GRP	CONTROL GRP.
Number of families	15	10
Sex of patient	10 male, 5 female	8 male, 2 female
Age	30.73 yrs.	31.03 yrs.
Educational level (1)	*a*: 0% *c*: 60% *b*: 13.33% *d*: 26.66%	*a*: 18.19% *c*: 36.36% *b*: 45.45% *d*: 0%
Years since first diagnosis	6.467	7.073
Months since last admission	22.545	18.53
Number of admissions	2.33	2.24
Course over last two years	9 re-admissions 6 changes of medication	8 re-admissions 2 changes of medication
Type of schizophrenia	11 paranoid 4 non-paranoid	6 paranoid 4 non-paranoid
Age at onset	24.267 yrs.	24.55 yrs.
Family size	2.867 members	3.55 members
Medication (mg chlorpromazine)	295.386	401.091
Type of family	15 parental	9 parental, 1 marital

* These data are as presented at the start of the programme.
(1): *a* = no education, *b* = primary, *c* = secondary, *d* = university.

(1989) Spanish version of the Family Stress Scale.

- Phase I (*Psychoeducational*). The family members in the experimental group, in groups of around five families, received information over 15 weekly 2´5 hour sessions about the illness (symptoms, etiology, treatment, etc.), about what the family can do (role of stress in the course of schizophrenia, how to cope with the illness, etc.) and about practical procedures related to the illness (relaxation, assertive behaviour, modification of irrational thoughts and problem-solving).
- Phase II (*Individual intervention*). Fifteen weekly sessions of 15 hours duration with each family (including the patient) individually. The aim was the application of what had been learned to concrete, everyday cases. In this phase a list was drawn up by consensus of aspects and behaviours that should be modified in order to improve family climate and reduce family stress, considering how to achieve such modification with the help of appropriate techniques (family counselling, behavioural contract, etc.).
- Phase III (*Follow-up*). This phase began after the previous phase in the experimental group and after selection in the control group, and lasted one year. It was carried out by means of telephone calls every 15 days, in which families reported on the state of the patient, taking of medication, hospital admissions and changes of medication. Moreover, experimental group families gave information on the use of the techniques learned. After the year of follow-up, a new evaluation of the control variables of the study was made (Muela, 1999; Muela and Godoy, in press).

A year after the follow-up of the Andalusia Study families, and no contact having been made with them in the interim, we proceeded to locate these families so that they could report on what had occurred during this second year.

Contact was made by telephone and by mail in all cases (some families requested this, and in other cases it was impossible to meet family members personally for a variety of reasons). The majority of the relatives that reported on the events during this second year of follow-up participated actively in the programme, though not all of them were those who had been interviewed at the beginning of it. Furthermore, in four cases (three of them in the control group) the family members had had no previous contact with the research team. Given this situation, it was decided not to evaluate the components of the programme (Expressed Emotion, Family Stress and level of knowledge about the illness) that had shown their change, after the first year of follow-up, in the experimental group.

Also, given that these relatives could only provide information on re-admissions (not being able to find or not knowing about reports on changes of medication), it was decided to adopt, as relapse criterion, re-admission of the patient, which had already been found to correlate significantly with relapse rate in the first year of follow-up.

Table 2
Relapse rate in second year of follow-up

		Relapses
Experimental group	0/15	(0%)
Control group	4/10	(40%)

RESULTS

During the second year of follow-up there were no re-admissions among the fifteen experimental group families. However, among the ten control group families there were four re-admissions (one of these four subjects was re-hospitalized on three different occasions). The differences are statistically significant according to Fisher's Exact Test (*p*=0.017). These results can be seen in Table 2.

DISCUSSION

Re-admission is not the most appropriate indicator for measuring relapse, so that the conclusions of this work should be accepted with caution. If we decided to adopt it as a definition of relapse it was because of the impossibility of obtaining reliable data on the other part of the relapse definition during the first year of follow-up: increase in medication due to exacerbation of symptoms. During the second year the periodical contact with families was lost, and while hospitalization is easily remembered, a change of medication may be forgotten if it is not preceded by important behavioural alterations in the patient (moreover, family members tend to keep admission reports, whilst changes of medication are usually reflected only on prescriptions, which are often thrown away). Therefore, admission was accepted as the best definition of relapse among those available.

Despite this shortcoming, it should be borne in mind that at the one-year follow-up of this same study the correlation between relapse and re-admission was statistically significant, as it was also in other studies, such as that of Brown, Birley and Wing (1972), up to the point that the results would be identical if, instead of the definition of relapse adopted by these authors, they had made their calculations using the re-admissions indicator. Moreover, in the two-year follow-up of the Salford Study (Tarrier, Barrowclough, Vaughn, Bamrah, Porceddu, Watts and Freeman, 1989), re-admission of patients was taken as the basis for the definition of relapse, since it was impossible to repeat with relatives the PSE ("Present State Examination"), the test with which relapse was defined during the first follow-up period.

Also, if we compare the relapse rates of the studies reporting on two-year follow-ups with those of the Andalusia Study, we find that there are no statistically significant differences between them. Thus, the Camberwell Study (Leff, Kuipers, Berkowitz, and Sturgeon, 1985) found, after two years of follow-up, relapse rates in the experimental and control groups of 14% and 78%, respectively. In the California Study (Falloon, Boyd, McGill, Williamson, Razani, Moss, Gilderman and Simson, 1985) the percentages were 17% and 83%, in the Pittsburgh Study (Hogarty, Anderson and Reiss, 1987), 32% and 66%, and finally, in the Salford Study (Tarrier, Barrowclough, Vaughn, Bamrah, Porceddu, Watts and Freeman, 1989), 33% and 59%. All of these differences are statistically significant within each study.

If we compare the data of all the experimental groups (including that of the Andalusia Study), we obtain a chi-squared of 7.78 with four degrees of freedom for a *p*=0.1. Comparison among control groups gives us a chi-squared of 6.604 with four degrees of freedom for a *p*=0.158. That is, there are no differences between the experimental groups of all the studies, nor between the control groups. This supports the choice of re-admission as relapse criterion. The data discussed can be seen in Table 3.

On not having contact with family members throughout the whole second year of follow-up, there is no guarantee of the strict criterion of patients' fulfilment of the medication regime that was adopted in the first year of follow-up. Nevertheless, relatives stated that the patients had not given up their medication at any time, or that, if they had, they had gone no longer than ten days without taking it (as far as they recalled).

It would have been interesting to check whether the reductions in Expressed Emotion and Family Stress and the increase in the family's knowledge about the illness achieved by the end of the first year of follow-up were maintained now, after the second year. Unfortunately, with the majority of the experimental group families and with all those of the control group it was only possible to make contact by telephone and mail, so that this form of contact was adopted in all cases (with three control group families and one experimental group family, the contact was even with a person who had had no direct relationship with the research team during the first year of follow-up).

On the other hand, given that the Andalusia Study clearly separates the intervention from the follow-up (which no other study does, since the follow-up period begins not at the end of the treatment but at the beginning of it), it was possible to compare the relapse rate of its experimental group over 19 months (the 12 of follow-up plus the seven for which the programme lasted) with those of the experimental groups of studies that have carried out two-year follow-ups (from the beginning of the respective programmes). The control group of the Andalusia Study had a one-year follow-up, and given that it did not carry out the programme, it is comparable with neither the 19-month follow-up of that study's experimental group nor with the two-year follow-up of the control groups of the other studies.

Thus, in the Andalusia Study, 27% of the patients in the experimental group relapse in this period (from the beginning of the programme until the end of the one-year follow-up), versus 32% in the Pittsburgh Study, 33% in the Salford Study, 14% in the Camberwell Study and 17% in the California Study. There are no statistically significant differences between any of the five groups (a chi-squared of 2.309 with four degrees of freedom for a *p*=0.679).

Taking into account the different forms of carrying out the follow-up in the different studies, Table 4 shows the percentage of relapses (at different points of the follow-up) for the experimental groups of each of the studies mentioned.

This indicates that the intervention programme used in the Andalusia Study is effective in the long term, to the same extent as those used in the other mentioned studies. Currently, the effectiveness of Family Intervention in schizophrenia is unquestionable. Bearing in mind the results of the five cited studies, it can be affirmed that the effects of these programmes are maintained for at least two years.

Table 3
Comparison of experimental and control groups of studies reporting two-year follow-ups

Relapses by groups					
Studies	Relapses Experimental Groups		Relapses Control Groups		*p*
Camberwell	1/7	(14%)	7/9	(78%)	0.02
California	3/18	(17%)	15/18	(83%)	<0.001
Pittsburgh	7/22	(32%)	23/35	(66%)	0.013
Salford	8/24	(33%)	17/29	(59%)	0.05
Andalusia	0/15	(0%)	4/10	(40%)	0.017
p	0.1		0.158		

Table 4
Percentage of relapses in the experimental groups of the different studies with different follow-ups (measured from start of study)

	9 MONTHS	12 MONTHS	19-24 MONTHS
CAMBERWELL	8%	---	14%
CALIFORNIA	6%	---	17%
PITTSBURGH	9%	19%	32%
SALFORD	12%	---	33%
ANDALUSIA	---	20%	27%

ACKNOWLEDGEMENTS

This work was supported in part by the grants from the Spanish Ministry of Education and Culture (DGICYT, PB 94-0820) and the Junta de Andalucía (CTS 0267).

REFERENCES

Álvarez, L. & Gutiérrez, E. (1989). Spanish versión of the Family Stress Scale (FSS) by Pasamamick et al., 1967. Unpublished material.

Anderson, C., Reiss, D. & Hogarty, G. (1986). *Esquizofrenia y familia.* Buenos Aires: Amorrurtu.

Barrowclough, C. & Tarrier, N. (1990). Social functioning in schizophrenic patients. I: The effects of expressed emotion and family intervention. *Social Psychiatry and Psychiatric Epidemiology, 25,* 125-129.

Brown, G., Birley, J. & Wing, J. (1972). Influence of family life on the course of schizophrenic disorder: A replication. *British Journal of Psychiatry, 121,* 241-258.

Falloon, I., Boyd, J. & McGill, C. (1984). *Family care of schizophrenia.* New York: Guilford Press.

Falloon, I., Boyd, J., McGill, C., Razani, J., Moss, H. & Gilderman, A. (1982). Family management in the prevention of exacerbations of schizophrenia. *The New England Journal of Medicine, 306,* 1437-1440.

Falloon, I., Boyd, J., McGill, C., Strang, J. & Moss, H. (1981). Family management training in the community care of schizophrenia. In M. Goldstein (Ed.). *New developments in interventions with families of schizophrenics.* San Francisco: Jossey-Bass.

Falloon, I., Boyd, J., McGill, C., Williamson, M., Razani, J., Moss, H., Gilderman, A. & Simson, G. (1985). Family management in the prevention of morbidity of schizophrenia: Clinical outcome of a two year longitudinal study. *Archives of General Psychiatry, 42,* 887-896.

Goldstein, M. & Kopeikin, H. (1981). Short and long term effects of combining drug and family therapy. In M. Goldstein (Ed.). *New developments in interventions with families of schizophrenics.* San Francisco: Jossey-Bass.

Goldstein, M., Rodnick, E., Evans, J., May, P. & Steinberg, M. (1978). Drug and family therapy in the aftercare of acute schizophrenics. *Archives of General Psychiatry, 35,* 1169-1177.

Gutiérrez, E. (1986). Spanish versión of the Camberwell Family Interview (CFI) by Vaughn & Leff, 1976. Unpublished material.

Hogarty, G., Anderson, C. & Reiss, D. (1987). Family psychoeducation, social skills training and medication in schizophrenia: The long and short of it. *Psychopharmacology Bulletin, 23,* 12-13.

Hogarty, G., Anderson, C., Reiss, D., Kornblith, S., Greenwald, D., Javna, C. & Madonia, M. (1986). Family psychoeducation, social skills training, and maintenance chemotherapy in the aftercare treatment of schizophrenia. *Archives of General Psychiatry, 43,* 633-642.

Leff, J., Kuipers, L., Berkowitz, R., Eberlein-Vries, R. & Sturgeon, D. (1982). A controlled trial of social intervention in the families of schizophrenic patients. *British Journal of Psychiatry, 141,* 121-134.

Leff, J., Kuipers, L., Berkowitz, R., Eberlein-Vries, R. & Sturgeon, D. (1983). Social intervention in the families of schizophrenics: Addendum. *British Journal of Psychiatry, 142,* 311.

Leff, J., Kuipers, L., Berkowitz, R., Eberlein-Vries, R. & Sturgeon, D. (1985). Controlled trial of social intervention in the families of schizophrenic patients. In M. Goldstein, I. Hand & K. Hahlweg (Eds.). *Treatment of schizophrenia: Family assessment and intervention.* New York: Springer.

Leff, J., Kuipers, L., Berkowitz, R. & Sturgeon, D. (1985). A controlled trial of social intervention in the families of schizophrenic patients: Two year follow-up. *British Journal of Psychiatry, 146,* 594-600.

Leff, J. & Vaughn, C. (1985). *Expressed emotion in families: Its significance for mental illness.* New York: Guilford Press.

Muela, J. (1999). *Programa de intervención familiar en esquizofrenia.* Jaén: Servicio de Publicaciones e intercambio científico de la Universidad de Jaén.

Muela, J. & Godoy, J. (in press). Program of familiar intervention in schizophrenia. Andalusia study.

Tarrier, N., Barrowclough, C., Vaughn, C., Bamrah, J., Porceddu, K., Watts, S. & Freeman, H. (1988). The community management of schizophrenia. A controlled trial of a behavioural intervention with families to reduce relapse. *British Journal of Psychiatry, 153,* 532-542.

Tarrier, N., Barrowclough, C., Vaughn, C., Bamrah, J., Porceddu, K., Watts, S. & Freeman, H. (1989). Community management of schizophrenia. A two year follow-up of a behavioural intervention with families. *British Journal of Psychiatry, 154,* 625-628.

Tarrier, N. & Birchwood, M. (1995). Hacer realidad el tratamiento comunitario de la esquizofrenia. In M. Birchwood & N. Tarrier (Coord.). *El tratamiento psicológico de la esquizofrenia.* Barcelona: Ariel.

Vaughn, C. & Leff, J. (1976). The influence of family and social factors on the course of psychiatric illness. *British Journal of Psychiatry, 129,* 125-137.

Zubin, J. & Spring, B. (1977). Vulnerability: A new view of schizophrenia. *Journal of Abnormal Psychology, 86,* 103-126.

Copyright 2002 by the
Colegio Oficial de Psicólogos. Spain

Psychology in Spain, 2002, Vol. 6. No 1, 61-67

ANTICIPATORY ANXIETY IN WOMEN RECALLED FOR FURTHER MAMMOGRAM BREAST CANCER SCREENING

Bonifacio Sandín, Paloma Chorot, Rosa M. Valiente, Lourdes Lostao* and Miguel A. Santed**
**Universidad Nacional de Educación a Distancia (UNED) and **Universidad Pública de Navarra*

En esta investigación se ha estudiado la ansiedad, el estado de ánimo deprimido, la depresión y la sintomatología somática asociados a la participación en pruebas complementarias de screening de cáncer de mama. Mil ciento noventa y cinco mujeres, con edades entre 45 y 65 años, fueron entrevistadas en dos momentos temporales (pre- y post-mamografía). Las participantes eran mujeres que acudían a realizar pruebas rutinarias de screening de cáncer de mama que fueron citadas para repetir las pruebas (i.e., pruebas complementarias) y mujeres que, asistiendo a las pruebas rutinarias, no fueron citadas a pruebas complementarias. Las variables fueron evaluadas a través de una escala de emociones tipo Likert de 10 puntos, el Cuestionario Estado-Rasgo (STAI) de Spielberger, Gorsuch y Lushene, la Escala de Síntomas Somáticos de Sandín y Chorot, y el Cuestionario de Depresión de Sandín y Valiente. Los resultados indican que las mujeres asistentes a las pruebas complementarias exhibían niveles significativamente más elevados de ansiedad anticipatoria y estado de ánimo deprimido que las mujeres que únicamente participaban en las pruebas rutinarias, durante los días que precedían a la realización de la prueba de mamografía. Los datos apoyan la hipótesis de que las mujeres que son llamadas para repetir la mamografía experimentan niveles entre moderados y altos de ansiedad y ánimo deprimido antes de la prueba, pero no sostienen la predicción de que tal efecto sea duradero (i.e., persista después de obtener el resultado negativo).

This investigation examined anxiety, depressive mood, depression and somatic symptoms associated with a second-stage screening for breast cancer. Interviews were conducted with 1195 women aged 45-65 in two time conditions (pre- and post-mammogram). Participants included women attending for routine breast cancer screening who were recalled for further mammogram, and women who were not recalled. Variables were assessed using a 10-point Likert emotion scale, the State-Trait Anxiety Inventory of Spielberger, Gorsuch and Lushene, the Somatic Symptoms Scale of Sandín and Chorot, and the Depression Questionnaire of Sandín and Valiente. Results indicated that women attending the second-stage screening exhibited significantly higher levels of anticipatory anxiety and depressive mood before the mammogram than women attending for routine screening. This emotional impact was not relevant two days after the mammogram. Data support the hypothesis that women recalled for further mammograms experience moderate to high levels of anxiety and depressive mood before the mammogram, though they do not sustain the prediction that this effect persists beyond receipt of the negative result.

Breast cancer is one of the commonest forms of cancer in women, and one of the major causes of death among them (Ascunce, 1991; Lostao, 1994). Nevertheless, it has been suggested that the death rate from breast cancer can be significantly reduced if the disease is detected early by means of mammogram screening programmes (Chamberlain & Palli, 1993). Increased acceptance of and participation in routine breast cancer detection screening (Ascunce & Del Moral, 1993; Gad & Rosselli del Turco, 1993; Rakowski, Stoddard, Rimer, Fox, Andersen, Urban, Lane, & Costanza, 1997) has led to an increase in the number of women that are recalled to repeat the mammogram (i.e., for a complementary check) due to some type of abnormality in the results of routine screening (suspicion of cancer risk, inconclusive data, problems of interpretation, etc.). The proportion of mammograms with abnormal or unsatisfactory results has been estimated at between 10% and 20% (Wardle & Pope, 1992), which means that the number of women affected by this type of problem is high.

Although only a small proportion of women who are required to have further tests turn out to have cancer, the fact of being recalled for a second screening may generate exceptional levels of psychological impact since, as some authors have pointed out (Skrabaneck, 1985), repetition of the mammogram represents a “false positive” (i.e., the woman temporarily experiences the diagnosis of cancer). Indeed, as Marteau (1994) has suggested, uncertainty about the result tends to be more worrying than actually receiving a negative or positive result.

The original Spanish version of this paper has been previously published in *Psicothema*, 2001, Vol. 13. No 2, 240-246

....

Correspondence concerning this article should be addressed to Bonifacio Sandín, Facultad de Psicología, Universidad Nacional de Educación a Distancia (UNED), 28040 Madrid (Spain).
E-mail: bsandin@psi.uned.es

Thus, given the high number of women whose results are inconclusive after the first mammogram, involvement in complementary tests may constitute in itself a considerable public health problem (in the USA, for example, over 4 million women a year are required to repeat the tests; Lerman, Trock, Rimer, Jepson, Brody & Boyce, 1991).

Despite the supposed relevance of studying the way women involved in such tests may be psychologically affected, few studies have dealt specifically with this issue, and the information available is scarcely conclusive. Ellman, Angelin, Christians, Moss, Chamberlain and Maguire (1989) found no significant differences in psychiatric morbidity between a group of women that attended routine screening and a group attending a second-stage screening, even though they did observe the latter to present higher levels of anxiety during the period just before the test. In any case, these differences in anxiety disappeared after three months of follow-up. Likewise, in a recent study, Clutton, Pakenham and Buckley (1999) found that women who, participating in a routine breast cancer screening programme, were required to repeat the mammogram, did not present significant levels of anxiety and psychopathology six weeks after the repetition.

Other studies, however, have found that women recalled for further tests suffered high psychological impact – and that this impact was greater than that experienced by women undergoing only the routine screening. This phenomenon has been observed both in the short term (i.e., during the few days just before the test; Cockburn, Staples, Hurley, De Luise, 1994) and in the longer term (three months after the second mammogram; Lerman et al., 1991). The latter group of authors concluded that the psychological effect (anxiety) and perception of risk persisted, despite the fact that the women had been adequately informed that they did not have cancer.

Apart from the fact that empirical evidence on the psychological effects associated with second-stage screening is scarce and inconclusive, the studies that have approached this phenomenon up to now have had some weaknesses, such as using fairly unrepresentative samples (sample size is generally less than 100, and/or women attending routine tests are not always used as control group) and retrospective or transversal methods. It would appear necessary, therefore, to carry out further research aimed specifically at examining the possible adverse psychological effects (i.e., the “psychological costs”) associated with participation in complementary breast cancer screening.

The aim of the present work was to study the anticipatory anxiety associated with complementary tests in women participating in the Breast Cancer Detection Programme (*Programa de Detección de Cáncer de Mama*, PDCM) run by the Government of Navarra in north-eastern Spain (Ascunce & Del Moral, 1993). According to our main hypothesis, women recalled for a second mammogram should experience higher levels of anticipatory anxiety and display other negative emotional features (depressive mood, depression and somatic symptoms) before the test than women attending a routine screening. In a second hypothesis we postulated that the effects associated with the second screening would not persist beyond receipt of a negative result. To test these hypotheses we used a large sample of women that attended the PDCM and a pre-post design.

METHOD

Design

Psychological assessments were carried out in two time conditions. (1) *Pre-mammography*: data were obtained immediately prior to the mammogram; (2) *Post-mammography*: data were obtained two days after the mammogram. Women were divided into two groups, a second-stage screening group (SSS group) and a routine screening group (RS group). Only the SSS group provided data in Time Condition 2.

Participants

We used a total sample of 1,200 women that had mammograms as part of the PDCM of Navarra. Two groups were formed: (a) one group of 600 women who were called back to repeat the screening, due to the fact that their routine mammogram had been inconclusive or shown some sign of abnormality [*second-stage screening group (SSS)*], and (b) a control group of 600 women who, having participated in the routine screening, were not recalled to repeat it [*routine screening group (RS)*]. The 600 subjects in the RS group were selected at random from a sample of 1,600 women that had just had routine breast cancer screening. Of the 1,200 women definitively selected (both groups), five were diagnosed with breast cancer and excluded from the study. The women’s age ranged from 45 to 65 years, and corresponded to the age range included in the PDCM. There were no differences between the two groups in age, educational level or marital status. Seventy-six percent of the women were aged 45 to 55, 59% had only elementary education and 76.5% were married.

Measures

Sociodemographic variables. Sociodemographic variables included age, educational level and marital status. They were assessed during the first interview by means of an individual interview sheet.

Self-rating variables. Self-rating was used to assess anticipatory anxiety ("To what extent have you felt anxious or nervous?") and depressive mood ("To what extent have you felt sad or depressed?") for the two days prior to the test, by means of an analog scale with points labelled from "None/absent" (1) to "A lot/Extremely severe" (10). This measure test is equivalent to that used by Lerman et al. (1991).

State-Trait Anxiety Inventory (STAI). We used the Spanish version adapted by Técnicos Especialistas Asociados (TEA; Spielberger, Gorsuch & Lushene, 1982), applying Part 1, which specifically assesses the state of anxiety. This part (STAI-S) comprises 20 items, which participants answered with an indication of how they felt during the two days prior to the screening, on a 5-point scale that went from "Not at all" (0) to "A lot" (4).

Somatic Symptoms Scale (*Escala de Síntomas Somáticos*, ESS) by Sandín and Chorot (Sandín, Valiente & Chorot, 1999). We used the short form, which includes 17 items related to bodily sensations that usually accompany states of anxiety and panic reactions, such as feelings of dizziness, tachycardia, irritability, vertigo, or muscular trembling. Subjects answered the questionnaire indicating the extent to which they had experienced each of the symptoms during the two previous days, on an intensity scale from "Not at all" (0) to "A lot" (4). Data supporting the reliability and validity of the questionnaire have been provided by its authors (Sandín & Chorot, 1991). Data have also been offered on the temporal consistency of the ESS, obtaining test-retest correlations, for periods of one month, ranging from 0.73 to 0.82 (Santed et al., 1994). Given that the present short form of the ESS had not been used previously, we calculated the Cronbach's alpha coefficients, which were found to be as follows: 0.84 (SSS group) and 0.81 (RS group) for the pre-mammogram situation, and 0.79 (SSS group) for the post-mammogram situation (2 days).

Depression Questionnaire (*Cuestionario de Depresión*, CD) by Sandín and Valiente (Sandín et al., 1999). This is a 16-item questionnaire designed for rapid assessment of clinical depression, as opposed to mere depressive mood. The questionnaire was constructed on the basis of symptoms required in the DSM-IV (APA, 1994) for the diagnosis of major depression. It includes mood symptoms (e.g., "I spent the day crying or wanting to cry"), cognitive symptoms (e.g., "I found it difficult to think or concentrate") and somatic symptoms (e.g., "I slept well"). Participants responded to the questionnaire indicating the extent to which they had experienced each of the symptoms during the previous two days, on a frequency scale ranging from "Never or hardly at all" (0) to "Almost all the time" (4). In the present study we obtained alpha coefficients ranging from 0.88 (in the SSS group) to 0.89 (in the RS group), indicating that the questionnaire has high internal consistency. The Depression Questionnaire has convergent and discriminant validity, since it correlated strongly with other depression measures, and moderately or weakly with constructs such as worry, fear or psychopathic thoughts (Sandín, Chorot, Lostao, Valiente, Buceta & Fernández-Soto, 2000).

Procedure

All the psychological measures were applied by means of individual interviews. During the pre-mammography period we assessed sociodemographic information and applied the psychological measures. During the post-mammography condition we repeated the psychological measures, though only for the women in the SSS group. All women were interviewed individually by a (female) psychologist, who administered the questionnaire packages. The interviews corresponding to the post-mammography condition were made by telephone, an appointment having previously been arranged between the interviewer and the participant.

All participants were informed of the results of the mammogram. The women required to take a second screening received the medical information just after the interview and the medical examination. In accordance with the PDCM procedure, the women taking the routine screening received notification of the medical results by post, within approximately 15 days, unless the mammogram showed inconclusive or abnormal results, in which case they were notified by telephone, in a standard format, that they were required to take complementary tests within two days. The types of information given about the results of the medical examination (benign pathologies) were as follows (we indicate first the number of subjects of the SSS group assigned to each category of information, and second, the number in the RS group): no pathology (471 and 509), liquid cysts (88 and 46), fatty cysts (18 and 6), nodules or fibroadenomas (13 and 30), and calcifications (7 and 7).

RESULTS

Mean values corresponding to the five psychological variables during the pre-mammography condition are shown in Table 1. For statistical analysis of differences between groups we carried out *t* tests for independent groups. As the table shows, there are significant differences between the groups for all the variables except somatic symptomatology (ESS) and depression (CD).

This suggests that psychological effects associated with the second-stage screening are related mainly to anticipatory anxiety prior to the medical examination, whether this anxiety is assessed by means of the STAI-S questionnaire or through self-rating. It can be observed, however, that the mean values of the SSS group are not high. Thus, for example, mean score in anxiety (self-rating) is 4.17, a value that indicates a moderate level of anticipatory anxiety. For the RS group, the anxiety level obtained by means of self-rating does not reach a value of 2, indicating that women who participated in the routine tests experienced low levels of anxiety. As regards anxiety measured by means of the STAI-S, given that the 20 items on this questionnaire were quantified using a 0-4 scale, a mean score of "moderate to severe anxiety" should attain at least a value of 40, a level considerable higher than those attained by either group of women.

As far as somatic symptomatology and depression are concerned, the mean scores indicate that neither those in the SSS group nor those in the RS group presented either somatic symptoms or depression associated with the breast cancer screening (the values for these two variables are low, and denote an absence of relevant somatic or depressive symptoms).

Figure 1

Percentage of women who presented during the pre-mammography condition scores from moderate to severe for the variables of anticipatory anxiety (anxiety by self-rating and STAI-S) and depressive mood. (SSS = second-stage screening, RS = routine screening).

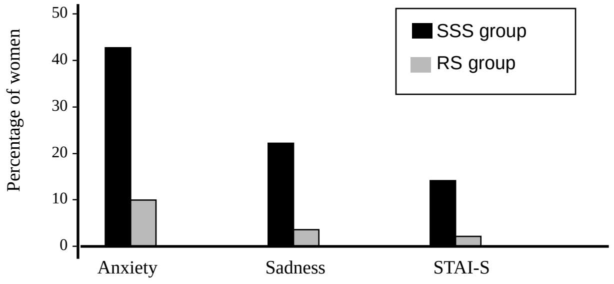


Figure 1 shows the percentages of women with scores from moderate to severe (scores between 5 and 10) in the self-rating variables (anxiety and depressive mood) and from "moderate" to "a lot" (scores between 40 and 80) in the STAI-S variable. As it can be seen, the frequencies, while not extremely high, are much greater in the three variables for the SSS group (χ^2 =154.32, *df* = 1, *P* < 0.001). Close to 50% of the women in the SSS group experienced significant levels of anxiety (self-rating), but only 11.2% of those in the RS group did so. Eighty-six (14.4%) of the women from the SSS group experienced between moderate and high anticipatory anxiety (STAI-S), but only 4 (0.7%) of those from the RS group.

The percentages of women from the SSS group that showed values from moderate to severe in these three psychological variables decreased drastically during the post-mammography condition (2 days later), with the figures as follows (number of subjects in brackets): 4% (24) for anxiety (self-report), 3.2% (19) for anxiety (STAI-S), and 3.7% (22) for depressive mood. As it can be seen, few women appear to experience anxiety and/or depressive mood during the days following the complementary mammography.

Finally, we were interested in determining pre-post differences in the values of these three variables for the women from the SSS group. Also, given that there were women who continued to experience significant levels of anxiety during the post-condition, it was important to study the possibility of its being related to the type of medical information provided to the women after the screening. For this purpose we carried out a repeated-measures analysis of variance (ANOVA) for each of the three dependent variables (i.e., anxiety by self-rating, anxiety by STAI-S, and depressive mood). The general ANOVA design used was a two-way (5x2) group (no

Table 1

Differences between groups in the pre-mammography condition

	SSS group		**RS group**			
	(*n* = 597)		**(*n* = 598)**			
Psychological variables	**Mean**	***SD***	**Mean**	***SD***	***t***	***p***
Anxiety (self-rating)	4.17	2.74	1.85	1.83	17.21	< 0.001
Depressive mood (self-rating)	2.82	2.50	1.83	0.92	14.52	< 0.001
State of anxiety (STAI-S)	29.18	8.29	19.87	5.30	23.05	< 0.001
Somatic symptoms (ESS)	5.01	6.04	4.20	5.11	1.74	ns
Depression (CD)	7.61	5.21	7.17	4.39	1.07	ns

Note: SSS = second-stage screening, RS = routine screening, ns = *P* > 0.05.

pathology vs. liquid cysts vs. fatty cysts vs. nodules or fibroadenomas vs. calcifications) x condition (pre- vs. post-mammography).

As Table 2 shows, mean values in anxiety (both measures) and depressive mood decrease drastically from the pre- to the post-condition. Moreover, taking into account these values, in general terms it can be affirmed that psychological effects seem to disappear completely two days after the second-stage screening and after receipt of the corresponding medical information. The ANOVAs carried out allowed us to confirm this phenomenon statistically, since the effect of the pre/post factor was significant for the three dependent variables studied [anxiety (self-rating): $F(1,592) = 73.7$, $P < 0.001$; anxiety (STAI-S): $F(1,592) = 1067.31$, $P < 0.001$; depressive mood: $F(1,592) = 14.29$, $P < 0.001$]. The effect of the group (type of information) factor was not found to be statistically significant for any of these three variables, despite the fact that women from the "calcifications" group presented higher mean levels than the rest of the women (see Table 2). Even though the interaction group x condition was not significant either, there was a trend towards statistical significance with regard to the two anxiety variables ($P < 0.09$ for anxiety by self-rating, $P < 0.06$ for anxiety by the STAI-S). This interactive trend suggests the existence of a low decrease in both forms of anxiety associated with positive report of calcifications.

DISCUSSION

The two main objectives of the present study were (1) to analyze the psychological effects or costs (basically anticipatory anxiety) for women attending a second-stage breast cancer screening, and (2) to determine whether this psychological impact disappeared within a short period of time (2 days after the examination). The data obtained indicate a significant level of psychological impact on anxiety, and a lower level on depressive mood. We did not find, however, a relevant effect on somatic symptomatology or depression (understood in terms of clinical depression, rather than mere sad or depressive mood).

With regard to the first issue, our data show that women in the second-stage screening (SSS) group experienced higher levels of anticipatory anxiety (evaluated both through self-rating and through the questionnaire) than the women in the routine screening (RS) group. These results, which suggest that there is greater psychological impact on women required to repeat the mammogram than on those not called back to repeat it, are congruent with those of Lerman et al. (1991) and Cockburn et al. (1994). In both studies the authors found significantly higher levels of psychological impact in women who were recalled for a second screening.

The absence of significant differences between the two groups of women in somatic symptomatology (ESS) and depression (CD) may be due to the fact that these varia-

Table 2
Means and *SD*s pre-/post-mammography in anxiety (self-rating and STAI-S) and depressive mood (self-report) for SSS (second-stage screening) group subjects according to type of information on pathology after mammography.

Variables according to type of information	Pre-mammography		Post-mammography	
	Mean	***DT***	**Mean**	***DT***
No pathology				
Anxiety (self-rating)	4.30	2.80	1.24	0.95
Depressive mood (self-rating)	2.92	2.58	1.23	1.02
State of anxiety (STAI-S)	29.46	8.55	4.37	9.72
Liquid cysts				
Anxiety (self-rating)	3.68	2.49	1.51	1.53
Depressive mood (self-rating)	2.60	2.24	1.49	1.49
State of anxiety (STAI-S)	28.07	7.47	6.48	13.94
Fatty cysts				
Anxiety (self-rating)	3.39	2.33	1.00	0.00
Depressive mood (self-rating)	1.89	1.53	1.00	0.00
State of anxiety (STAI-S)	27.61	7.06	3.28	6.44
Nodules/fibroadenomas				
Anxiety (self-rating)	3.46	2.44	1.23	0.60
Depressive mood (self-rating)	1.85	1.21	1.08	0.28
State of anxiety (STAI-S)	26.69	6.94	4.00	8.35
Calcifications				
Anxiety (self-rating)	5.00	3.37	2.43	2.99
Depressive mood (self-rating)	3.24	3.18	2.57	3.05
State of anxiety (STAI-S)	35.57	9.69	13.29	13.35

bles have greater psychopathological and clinical significance than the rest of the variables studied. The ESS (Somatic Symptoms Scale) questionnaire evaluates symptoms of autonomic activation and psychophysiological tension related to panic reactions and generalized anxiety disorder. The CD (Depression Questionnaire) provides a measure of depression, especially clinical depression. In our view, given that the extent of the emotional impact associated with second-stage screening is not high, we may expect: (1) not to find high scores in somatic symptoms and clinical depression, and (2) not to find significant differences between the two groups of women in these two variables. These results support the preliminary data published by Ellman et al. (1989), who found no significant differences in psychiatric morbidity (though they did find significant differences in anxiety) between women that participated in routine tests and those attending second-stage screening.

As we had expected, the psychological effects associated with second-stage screening seemed to disappear a short time after having taken the second screening and received reassuring information (i.e., information that there is no cancer or high risk of cancer). These results support the idea that the psychological effect related to this screening is a phenomenon of moderate intensity and short duration. Results in a similar direction were published recently by Clutton et al. (1999), who found that the psychological effects produced by second-stage breast cancer screening were practically irrelevant six weeks after the examination. Nevertheless, our data are in contrast to those of Lerman et al. (1991), who found that both anxiety and worry remained relatively high three months after the repeated mammogram. A problem with this last study, however, is that the assessment was carried out on a single occasion, and retrospectively.

The results of the present work suggest the existence of some emotional impact in women recalled for further mammography. The effects are not serious and are of short duration, since they are circumscribed to variables of emotional state and the time period between notification of the second appointment and receipt of information on the results of the second-stage screening. It seems clear, however, that a considerable percentage of the women who take such screening present high levels of anxiety and depressive mood (around 50% of such women in our sample experienced moderate to high levels of anxiety during the two days prior to the screening). Thus, although our data do not confirm the suggestion of some authors (Schmidt, 1990; Skrabaneck, 1985) that breast cancer screening, and in particular second-stage screening, may constitute – on generating the expectation of "false positives" – a potent source of serious psychological disturbance, it is nevertheless true that such screening may to some extent reduce women's quality of life, and at least affect some variables of emotional well-being.

If we take as a point of reference Marteau's (1994) emotion-cognition model, according to which mammography is associated with high uncertainty in women from the non-clinical population, it can be suggested that the second-stage screening constitutes a paradigmatic situation of the generation of uncertainty, and consequently of anticipatory anxiety and other possible negative emotion al responses. Given that, according to our data, the effects tend to be of short duration, remaining for the time that elapses between notification of the second appointment and receipt of the reassuring results, an important implication of the present work is that the best way of reducing the emotional impact on women would be to shorten this period as much as possible. It also emerges from the present study that the general strategy established by the PDCM team in Navarra is appropriate. Giving the reassuring information (i.e., the negative result) to the woman as soon as possible is an adequate method of reducing the negative effects. Anxiety levels might perhaps also be reduced by providing the woman, on notification of the second appointment, with some type of complementary information that would lower the level of uncertainty of the second screening. Such information might consist in informing her of the remoteness of the possibility of cancer or cancer risk, adducing technical reasons, including specific instructions, etc.

Finally, we have found that, even though the information was not sufficiently conclusive due to the small number of women affected (seven in the SSS group), women with calcifications tend to present high levels of negative impact during the two periods (pre- and post-). It is therefore perhaps necessary to provide such women with some kind of additional information – medical or psychological –, with the object of reducing their level of distress (and possibly of uncertainty), in addition to studying this benign pathology in more depth.

To the best of our knowledge, this is the first work focusing specifically on the study of psychological effects associated with second-stage breast cancer screening based on a representative sample of the population, employing temporal measures and using a control group of women who participated in routine screening. The results of our study serve to clarify some of the contradictory aspects referred to in the literature on the issue in question. Given that the emotional effects appear to be greater for some women than for others, it would be important to analyze in future research the possible role of dispositional variables (e.g., neuroticism, attitudes towards disease) and social variables (e.g., social support) in this phenomenon.

A possible limitation of the present study resides in the fact that the interviews were applied face-to-face during the pre-mammography condition and by telephone during the post-mammography condition. Nevertheless, the possible effect of this on the results is minimal, since the interviews were conducted by the same psychologist and in a similar way in each case. Moreover, there is evidence that the telephone interview provides data as valid and reliable as that provided by the traditional interview (Zapka, Bigelow, Hurley, Ford, Egelhofer, Cloud & Sachsse, 1996).

ACKNOWLEDGEMENTS

The present study was supported by research grant nº 2204, *Resolución* 634/98, from the *Departamento de Salud del Gobierno de Navarra*. We should like to thank Dra. Nieves Ascunce and Dr. Ángel Del Moral for their help and suggestions.

REFERENCES

APA (American Psychiatric Association) (1994). *Diagnostic and statistical manual of mental disorders, 4th edition (DSM-IV)*. Washington, DC: Author.

Ascunce, N. (1991). Mortalidad por cáncer en España. Movimiento natural de la población española. Prevención secundaria del cáncer. In J. Estapé & I. Burgos (Eds.), *Medicina General y Cáncer* (pp. 67-75). Barcelona: Doyma.

Ascunce, N., & Del Moral, A. (1993). Compliance in breast cancer screening: The Spanish experience. In A. Gad & M. Rosselli del Turco (Eds.), *Breast cancer screening in Europe* (pp. 47-53). Berlin: Springer-Verlag.

Chamberlain, J., & Palli, D. (1993). Evidence of the effectiveness of breast cancer screening. In A. Gad & M. Rosselli del Turco (Eds.), *Breast cancer screening in Europe* (pp. 15-23). Berlin: Springer-Verlag.

Clutton, S., Pakenham, K.I., & Buckley, B. (1999). Predictors of emotional well-being following a “false positive” breast cancer screening result. *Psychology and Health, 14,* 263-275.

Cockburn, J., Staples, M., Hurley, S., & De Luise, T. (1994). Psychological consequences of screening mammography. *Journal of Medical Screening, 1,* 7-12.

Ellman, R., Angelin, N., Christians, A., Moss, S., Chamberlain, J., & Maguire, P. (1989). Psychiatric morbidity associated with screening for breast cancer. *British Journal of Cancer, 60,* 781-784.

Gad, A., Rosselli del Turco, M.(1993). *Breast cancer screening in Europe*. Berlin: Springer-Verlag.

Lerman, C., Trock, B., Rimer, B.K., Jepson, C., Brody, D., & Boyce, A. (1991). Psychological side effects of breast cancer screening. *Health Psychology, 10,* 259-267.

Lostao, L. (1994). *Factores psicosociales asociados a la participación en un programa de prevención de cáncer de mama.* Madrid: UNED (doctoral thesis).

Marteau, T.M. (1994). Psychology and screening: Narrowing the gap between efficacy and effectiveness. *British Journal of Clinical Psychology, 33,* 1-10.

Rakowski, W., Stoddard, A.M., Rimer, B.K., Fox, A., Andersen, M.R., Urban, N., Lane, D.S., Costanza, M.E. (1997). Confirmatory analysis of opinions regarding the pros and cons of mammography. *Health Psychology, 16,* 433-441.

Sandín, B., & Chorot, P. (1991). *Escala de Síntomas Somáticos (ESS)*. Poster presented at III Congress of Psychological Assessment. Barcelona, 25-28 September.

Sandín, B., Chorot, P., Lostao, L., Valiente, R.M., Buceta, J.M., & Fernández-Soto, O. (2000). *Afectación psicológica en mujeres que acuden a pruebas complementarias en un programa de screening de cáncer de mama.* Madrid: UNED (unpublished).

Sandín, B., Valiente, R.M., & Chorot, P. (1999). Evaluación del estrés psicosocial. In B. Sandín (Ed.), *El estrés psicosocial: Conceptos y consecuencias clínicas* (pp. 245-316). Madrid: Klinik.

Santed, M. A., Chorot, P., Sandín, B., Jiménez, M. P., & García-Campayo, J. (1994). *Stress assessment: A new microevents scale*. Poster presented at 23rd International Congress of Applied Psychology. Madrid, 17-22 July.

Schmidt, J.G. (1990). The epidemiology of mass breast cancer screening - A plea for a valid measure of benefit. *Journal of Clinical Epidemiology, 43*, 215-225.

Skrabanek, P. (1985). False premises and false promises of breast cancer screening. *Lancet,* 316-320.

Spielberger, C.D., Gorsuch, R.R., & Lushene, R.E. (1982). *STAI. Cuestionario de Anxiety Estado/Rasgo.* Madrid: TEA.

Zapka, J.G., Bigelow, C., Hurley, T., Ford, L.D., Egelhofer, J., Cloud, W.M., & Sachsse, E. (1996). Mammography use among sociodemographically diverse women: The accuracy of self-report. *American Journal of Public Health, 86,* 1016-1021.

Wardle, J., & Pope, R. (1992). The psychological costs of screening for cancer. *Journal of Psychosomatic Research, 36,* 609-624.

Copyright 2002 by the
Colegio Oficial de Psicólogos. Spain

Psychology in Spain, 2002, Vol. 6. No 1, 68-76

QUALITY OF LIFE PARAMETERS IN TERMINAL ONCOLOGICAL PATIENTS IN A HOME CARE UNIT

Ana González, Concepción Fernández, Gerardo García*, Jorge Soler, Clavel Arce* y José Cueto
*Universidad del Oviedo y * Hospital de Cabueñes (Gijón)*

El estudio pretende revisar la metodología y las dificultades asociadas a la evaluación de la calidad de vida en pacientes oncológicos en situación terminal y avanzada. Para ello se evalúa la calidad de vida informada a través de una escala estandarizada, el cuestionario de calidad de vida de la EORTC, el QLQ-C30, y se observa si es sensible a la recepción de cuidados paliativos en el domicilio. Se pretende al mismo tiempo, comparar estos resultados con los posibles cambios reportados por los pacientes sobre sus síntomas físicos, sobre la percepción del apoyo social y sobre los marcadores de ansiedad depresión, tras el ingreso en una Unidad de Hospitalización a Domicilio (UHaD). El trabajo se realizó con 42 pacientes a los que se evaluó en dos ocasiones, al ingreso en la UHaD y al cabo de 1, 2, 3 ó 4 semanas. Se recogen variables sociodemográficas y clínicas, y los datos se analizan mediante análisis de covarianza de medidas repetidas, y a través de t de Student para muestras relacionadas. Los resultados indican que tan sólo la variable dolor del QLQ-C30 cambia significativamente tras el ingreso en la UHaD (p<.045). Por el contrario, la mayoría de las variables referidas a sintomatología clínica y evaluadas por el médico, descienden significativamente tras el ingreso en la unidad (vómitos: p<.003; dolor: p<.000; estreñimiento: p<.000 y sueño: p<.000). Los niveles de ansiedad y depresión, que al ingreso ya revelan ausencia de patología, no se modifican desde los valores iniciales. El índice de Karnofsky no cambia significativamente tras el ingreso en la Unidad. Se discute la adecuación del uso del término calidad de vida en la enfermedad terminal.

In this study the quality of life of 42 terminal oncological patients within a home care unit was evaluated using the questionnaire QLQ-C30 and self-report measures. Patients were evaluated at two time points, when they entered in the Unit and after one, two, three or four weeks. Clinical and demographic variables were registered and analyzed using repeated-measures covariance analyses and the Student t-test. Results revealed that only pain evaluated through QLQ-C30 changed significantly after admission to the Unit (p<.45). On the other hand, most of the clinical symptomatology variables, evaluated through the SCS, showed significant decreases (vomiting, p<.003; pain, p<.000; constipation, p<.000; sleep, p<.000). Anxiety and depression levels, which were below clinical levels on admission to the Unit, showed no change at the second evaluation. Nor did the Karnofsky Index show significant changes. The quality of life concept in relation to terminal illness is discussed.

A growing interest in the measurement of the "quality of life" concept has given rise to numerous interpretations and some confusion with respect to its definition and objectivity. Quality of life indicators in oncology have ranged from the purely physiological and physical to complex questionnaires based on the psychological repercussions of the illness and the social activities of patients. Some of the first measures used were designed to quantify patients' "health status" from perspectives such as perceived distress (Hunt, McEwen and McKenna, 1985), impact of the illness (Bergner, Bobbitt and Pollard, 1976) physical functioning (Mahoney and Barthel, 1965) and degree of patient satisfaction (Lough, Lindsay, Shinn and Stotts, 1985). Thus, quality of life was evaluated using a variety of indicators, many of which are informative with respect to the life led by patients but not with respect to its quality. In any case, many of the instruments considered appropriate for the assessment of this concept have been applied in the early phases of illness, and there are still no instruments validated with Spanish samples to evaluate quality of life during the terminal stages of illness (Pratheepawanit, Salek and Finlay, 1999; Padierna and Fernández, 2001). As Tierney, Horton, Hannan and Tierney (1999) remind us, studies on the quality of life in terminal patients have basically concentrated on the relief of symptoms, rather than on a formal assessment of patient satisfaction. Furthermore, the evaluations

---

The original Spanish version of this paper has been previously published in *Psicothema*, 2001, Vol. 13. Nº 2, 310-317

....

Correspondence concerning this article should be addressed to Ana González Menéndez. Facultad de Psicología. Universidad de Oviedo, 33003 Oviedo. Spain. E-mail: anagon@correo.uniovi.es

have more often been carried out by carers than by the patients themselves, due to the belief that the latter were too ill to provide valid information (Carnike and Carey, 1999). Subsequent studies that show these patients to be capable of giving information about their status indicate a low degree of correlation between their assessments and those of their carers (Keizer, Kozak and Scott, 1992; McMillan and Mahon, 1994). The consequence of this and other confusing aspects is that it is often impossible to identify exactly what is being measured and, furthermore, the basic principles of application of some specific measures are not clear.

One of the aims of this study is to review the methodology employed and the difficulties involved in evaluating *quality of life* in terminal cancer patients. For this purpose, the aim is to examine the usefulness of information provided by the QLQ-C30, one of the instruments most commonly used for assessing quality of life in terminal cancer patients, with respect to palliative care in the home. Another objective is to compare these results with possible changes in patients' reports about their physical symptoms and social support, and in anxiety and depression indicators after admission to a home care unit (HCU). Lastly, the suitability of the term *quality of* life in terminal illnesses is discussed, with a view to proposing a different and perhaps more pertinent concept in relation to this issue, that of "degree of comfort", or even " quality of death".

METHOD

Subjects

The sample was made up of 42 oncology patients (12 women and 30 men) with different types of terminal cancer at an advanced stage. Participants' age ranged from 46 to 90, with a mean of 69. In order to be included in the study, participants had to meet the following conditions: Karnofsky Index score ≥ 40, absence of cognitive dysfunction and the verbal consent of both patient and family to participate in the study. All participants were selected from the various departments (internal medicine, medical oncology, urology, emergency and others) within the HCU of the Hospital de Cabueñes (Gijón, northern Spain).

Material

Hospital Anxiety and Depression Scale (HAD, Zigmond and Snaith, 1983). The HAD was designed to evaluate the emotional state of subjects receiving non-psychiatric hospital outpatient attention. With the aim of avoiding false positives in the psychopathological assessment in these contexts, the authors excluded references to physical symptoms. The scale consists of 14 items divided into two subscales of anxiety and depression, each with 7 items. The content of the items refers to the patient's subjective perception in relation to psychological aspects associated with depression and anxiety disorders. Each item has 4 response alternatives rated according to a Likert scale with scores ranging from 1 to 4 measuring the intensity of perceived discomfort on the part of the patient. The authors situated the cut-off zone for the two subscales at between 8 and 10, considering cases for values over 11. Validity coefficient is 0.70, and with respect to reliability, Cronbach's alpha coefficient is 0.80 for each of the subscales.

Functional Social Support Scale (Duke-UNC, Broadhead, Gehlbach, De Gruy, Kaplan, 1988). This is a self-assessment scale made up of 11 items that record people's opinions on the availability of others capable of offering support in times of difficulty, on access to social relationships and on their own possibilities for empathic and emotional communication. This questionnaire evaluates two dimensions of functional social support; *confidant* (items 7, 8, 6, 4, 1 and 10; defined by the possibility of having access to people with whom problems can be discussed) and *affection* (items 11, 9, 2, 3 and 5; defined by the degree of access to people that provide affection). The response to each of these items is evaluated by means of Likert scale of 1 to 5 points. There is a Spanish version of the questionnaire with a reliability of 0.80.

Quality of Life Questionnaire for Cancer (QLQ-C30) European Organisation for Research and Treatment of Cancer (EORTC) (Aaronson et al., 1993; Sprangers, Cull, Bjordal, Groenvold and Aaronson, 1993). This instrument consists of 30 items distributed across 5 functional scales (physical functioning; role; social functioning; emotional functioning and cognitive functioning), and three symptom scales (fatigue, pain and nausea-vomiting). Likewise, the questionnaire incorporates a global health/quality of life scale and some individual items that evaluate different symptoms of the illness and/or treatment (dyspnea, insomnia, loss of appetite, constipation, diarrhoea and financial impact). The questions refer to a time period of one week and use a Likert-type response format (an example of this questionnaire can be seen in Appendix 1) .

Symptom Control Sheet (SCS). This is a register of symptoms designed and applied by the doctors at the HCU. It records some of the commonest and most incapacitating symptoms of terminal cancer (activity or mobility level, vomiting, dyspnea, pain, constipation and sleep). See Appendix 2 to consult this register and its correction.

Karnofsky Index. (Karnofsky and Burchenal, 1949). Used to evaluate the functional status of patients on a scale of 0-100 (0=death; 100=normal development).

PROCEDURE

All patients were assessed on two occasions – on admission to the HCU and after 1, 2, 3 or 4 weeks after admission. Given that the mean stay of terminal cancer patients in the HCU is around 4 weeks (García, Cueto, Arce and González, 1995), the second evaluation had to be carried out within 30 days of admission. The choice of date for this second evaluation was made on the basis of a random numbers table designed for this purpose. Group size according to time was as follows: 17 patients were re-assessed one week after admission to the HCU, 8 of the 42 were re-assessed after 2 weeks in the HCU, 9 patients 3 weeks after admission and 8 patients after 4 weeks in the HCU.

During the first interview with the subjects, in which they were asked for their consent to participate in the study, they were informed about its objectives and general structure. The verbal consent of patients and their relatives was requested by the doctor, who also introduced both patient and family to the person responsible for applying the psychological and quality of life instruments used. During this initial interview, which took place in all cases on the first day of admission to the HCU, the doctor in charge carried out the first measurement for the SCS, assessed the patient on the Karnofsky Index and determined the palliative treatment appropriate for the case. The HAD, DUKE-UNC and QLQ-C30 were applied by the psychologist on the day the patient was admitted. One, 2, 3 or 4 weeks after admission the patients were re-assessed following the same procedure. Three HCU doctors and a psychologist participated in the study. The latter was trained in and familiar with the application of these tests to terminal cancer patients. Total duration of the study was 18 months, and it was carried out in accordance with the availability of participating patients.

DATA ANALYSIS

A repeated-measures covariance analysis was carried out, in which the covariant was initial score on each of the scales and subscales applied on the day of admission to the HCU. Subjects' scores on each of the scales and subscales on entering the HCU were taken as the WITHIN variable and time elapsed between the date of admission and the second evaluation as the BETWEEN variable (1, 2, 3 or 4 weeks). This analysis was carried out with the aim of detecting changes in the variables evaluated as a function of the time of application of the second evaluation (1, 2, 3 or 4 weeks after admission).

Furthermore, and with the aim of checking for the presence or absence of changes after admission to the HCU in the variables used, the data were analyzed by means of the Student t-test for related samples.

RESULTS

The results are obtained from the total number of patients making up the sample (n=42). The variables considered were: age, indicators of anxiety and depression, perceived social support and the QLQ-C30 with its respective subscales. All those variables controlled by the doctor (SCS) were taken into account: activity level, vomiting, dyspnea, pain, constipation, sleep and Karnofsky Index.

Age

Mean age of the total sample (n=42) was 69.8 years, with a range of 46 to 90 years. By gender, men (n=30) had a mean age of 71.3 years and a range of 48 to 90, and the 12 women in the sample had a mean age of 66.16 and a range of 46 to 83.

HAD-A (Anxiety indicators evaluated by the HAD)

Raw score of the total study sample in the anxiety variable evaluated by the HAD was 7.97 from a maximum possible score of 21. By gender, the women had slightly higher anxiety scores than the men (9.08 and 7.53 for women and men, respectively). This difference was not statistically significant ($p<.409$).

The results indicate the absence of statistically significant differences between the two times of evaluation of the anxiety variable by the HAD ($p<.704$). Also, the repeated-measures covariance analysis indicates that regardless of the timing of application of the second evaluation (after 1, 2, 3 or 4 weeks), differences are not observed with respect to the anxiety levels of the subjects [$F= 1.6654$; $p<.225$].

HAD-D (Depression indicators evaluated by the HAD)

Mean level of depression for the total sample of the study was 7.30 out of a maximum possible score of 21. As occurred for the anxiety variable, the women presented higher levels of depression than the men (10.08 and 6.20 for women and men, respectively), this difference being statistically significant ($p<.001$).

On the other hand, despite the fact that a slight increase in the depression indicators was observed after the period spent in the HCU, these differences were not statistically significant (p<.196). The repeated-measures ANCOVA indicates the absence of significant differences between the different time points of evaluation of this variable.

DUKE-UNC

Perceived social support reported by the total of patients was 50.39 out of a maximum of 55 points. By gender, the women reported a lower degree of social support (48.08) compared to the 51.20 reported by the men. The differences observed were statistically significant (p<.012).

Perceived social support values among the 42 patients in the study did not show significant changes from the initial values at admission to the HCU (p<.714). (see Table 3). The repeated-measures ANCOVA indicated no significant differences between the two evaluation time points for this variable.

QLQ-C30

Detailed information on the mean scores obtained by subjects in the different functional subscales and individual items of the QLQ-C30 can be seen in Table 1. A model of this instrument is shown in Appendix 1.

By gender, the women presented global quality of life levels of approximately 3 points, while the men scored slightly higher (3.51). The difference by gender in this variable was also significant (p<0.54).

After the second evaluation, none of the five functional scales or the other symptom scales or individual items changed after admission to the HCU. Decreases were only detected with respect to the initial values on the scale referred to as *pain* and the individual item *insomnia*, though only those for pain reached statistical significance (p<.045 and p<.062, pain and insomnia, respectively) (see Table 3). The repeated-measures ANCOVA was not significant for any of the subscales or individual items of the QLQ-C30.

SCS-ACTIVITY

With respect to activity or mobility level, the total sample presented a mean activity of 2.07.

The range for this variable was established between 0 (goes out, total autonomy) and 4 (bedridden more than 80% of day) (see Appendix 4 and Table 2).

The variable activity or capacity of movement on the part of the patient also decreases in the second evaluation, though this decrease does not reach statistical significance (p<.065) (see Table 3). The repeated-measures ANCOVA indicates no statistically significant differences between the different evaluation time points for this variable.

SCS- VOMITING

With respect to vomiting reported by the patients before admission to the HCU, the total study sample shows a mean level of 0.80. See Appendix 2.

Vomiting reported on admission decreases significantly in the second evaluation (p<.003) (see Table 3). Statistically significant differences were not detected by the repeated-measures ANCOVA.

SCS-DYSPNEA

With reference to the dyspnea reported by the 42 patients in the study, most are situated at around 0.61, and the range of scores is from 0 (absence of dyspnea) to 4 (inca-

Table 1
Mean values obtained for the total sample in the QLQ-C30

VARIABLES	**SCORES** **Means/d**	**Range**
PHYSICAL FUNCTIONING	1.53/.27	1-2
ROLE	1.46/.17	1-2
SOCIAL FUNCTIONING	1.72/.89	1-4
EMOTIONAL FUNCTIONING	2.25/.77	1-4
COGNITIVE FUNCTIONING	1.69/.72	1-4
FATIGUE	2.89/.79	1-4
PAIN	2.27/1.04	1-4
NAUSEA/VOMITIN	1.79/.91	1-4
DYSPNEA	1.95/1.12	1-4
INSOMNIA	2.14/1.07	1-4
APPETITE	2.90/1.00	1-4
CONSTIPATION	2.61/1.26	1-4
DIARRHOEA	1.19/.59	1-4
GLOBAL SCALE	3.36/.81	1-7

Table 2
Mean values obtained by the total sample in the SCS

VARIABLES	**SCORES** **Means/d**	**Range**
ACTIVITY	2.09/1.12	0-4
VOMITING	0.80/1.27	0-4
DYSPNEA	0.61/1.01	0-4
PAIN	2.11/1.36	0-4
CONSTIPATION	0.61/.49	0-1
SLEEP	1.16/.79	0-2

pacitating dyspnea). See Appendix 2 and Table 3.

Although dyspnea levels fall with respect to the initial values, this decrease is not statistically significant (p<.071) (see Table 3). The repeated-measures ANCOVA was not significant for the dyspnea variable evaluated by means of the SCS.

SCS – PAIN

Mean pain level reported by the total study sample was 2.11 from a maximum of 4. The scoring range was from 0 (absent) to 4 (incapacitating). See Appendix 2 and Table 2.

After the second evaluation, pain decreased significantly with respect to the initial values (p<.000) (Table 3). The repeated-measures ANCOVA indicates the absence of statistically significant differences for the different evaluation time points in the SCS pain variable.

SCS-CONSTIPATION

On admission, the total sample reported constipation levels of 0.61. This variable ranged from 0 (absence) to 1 (constipation). See Appendix 2 and Table 2.

Table 3
Pre- and post-evaluation means and differences in the variables assessed

INSTRUMENT	PRE MEAN	POST MEAN	DIFFERENCES
HAD			
Anxiety	7.97	7.69	P<.704
Depression	7.30	8.40	P<.196
Duke-UNC	50.39	49.97	P<.714
QLQ-C30			
Physical Functioning	1.53	1.49	P<.281
Role	1.46	1.51	P<.210
Social Functioning	1.72	1.75	P<.863
Emotional Functioning	2.25	2.14	P<.477
Cognitive Functioning	1.69	1.67	P<.877
Fatigue	2.89	2.84	P<.943
Pain	2.27	1.91	P<**.045**
Nausea/vomiting	1.79	1.71	P<.565
Dyspnea	1.95	1.73	P<.277
Insomnia	2.14	1.76	P<.062
Appetite	2.90	2.80	P<.643
Constipation	2.61	2.42	P<.411
Diarrhoea	1.19	1.11	P<.519
Financial Functioning	1.33	1.19	P<.160
Global Scale	3.36	3.30	P<.761
SCS			
Activity	2.09	1.63	P<0.65
Vomiting	0.80	0.21	**P<.003**
Dyspnea	0.61	0.40	P<.071
Pain	2.11	0.66	**P<.000**
Constipation	0.61	0.23	**P<.000**
Sleep	1.16	0.33	**P<.000**

A statistically significant decrease with respect to the initial values was also observed in the constipation variable of the SCS (p<.000) (Table 3). The repeated-measures ANCOVA indicated the absence of statistically significant differences in the different evaluation times for this variable.

SCS- SLEEP

An initial score of 1.16 was found in the sleep variable evaluated by the doctor. Possible scores were 0 (well), 1(reasonably) and 2 (badly). See Appendix 2 and Table 2.

A statistically significant decrease with respect to the initial values was also observed in this variable (p<.000) (Table 3). The repeated-measures ANCOVA indicated the absence of statistically significant differences in the different evaluation times for this variable.

KARNOFSKY INDEX

The total sample presented a mean Karnofsky Index value of 52.8 and a range of 40 to 70. No significant changes were observed in the Karnofsky values after the second evaluation in the total sample of patients.

DISCUSSION

One of the objectives of this project involved identifying the demographic characteristics of the study sample with regard to the variables relevant to the terminal process of the illness. Some of these variables, such as the indicators of *anxiety and depression* and the *social support* available to the patient, were assessed using two psychological instruments, the HAD scale (Zigmond and Snaith, 1983) and the DUKE-UNC social support questionnaire (Broadhead et al., 1988) whose utility in the psychological evaluation of physical patients is widely documented in the specialized literature (De la Revilla, Bailón, de Dios, Delgado, Prados and Fleitas, 1991; Caro and Ibañez, 1992; Bredart et al., 1999 Skarstein, Ass, Fossa, Skovlund and Dahl, 2000). Another variable of growing interest in current oncological research is that of *quality of life* with reference to the terminal process of cancer. Despite a lack of consensus upon its definition, quality of life is generally understood as a multidimensional concept that encompasses psychological and social aspects, symptoms generated by both the illness and its treatment and the patient's level of functioning. Although since 1993 improving quality of life has been a priority objective in dealing with patients in the terminal stage of illness, there is still no valid and reliable instrument available for its evaluation within the context of palliative care (Rees, Hardy, Ling, Broadley, A'Hern, 1998). In order to assess quality of life we applied the EORTC questionnaire QLQ-C30, one of those most commonly used in oncology patients (Arrarás, Illaramendi and Valverdi, 1995). Finally, we consider and assess some of the variables related to the physical symptoms frequently presented by oncology patients in advanced stages of terminal illness (pain, dyspnea, vomiting, constipation and sleep) admitted to the HCU (García et al., 1995).

First of all, and with regard to the evaluation of anxiety and depression levels by means of the HAD scale, the results indicate that the sample of our study is halfway between the category "absence of symptoms" (scores of between 0 and 7 for both anxiety and depression) and "doubtful case" (scores of between 8 and 10 for both anxiety and depression), according to the criteria of the authors themselves (Zigmond and Snaith, 1983). In fact, the criteria proposed by Snaith (1983) indicate that for the inclusion or exclusion of a patient from the group with emotional disorders, only scores of between 11 and 21 would be clear indicators of anxiety and depressive disorders. If we bear in mind that, on admission to the HCU, the sample of this study presented mean levels of 7.97 and 7.30 for anxiety and depression, respectively, we can affirm that in our study the cancer patients, despite being in the so-called terminal phase of the illness, are psychologically well adjusted. This finding casts doubt on reports of emotional problems in patients with physical illnesses, and supports the idea that while oncology patients may present symptoms, especially of anxiety and depression, these symptoms should not necessarily be considered as constituting a clinical syndrome as the term is traditionally understood. It is also worth mentioning the differences found between men and women with respect to the evaluation of the two variables. In accordance with published findings over many years (Taylor, 1953; Zung, 1973; Spielberger, 1977; Zigmond and Snaith, 1983; Bredert et al., 1999), anxiety and depression are influenced by socio-demographic factors such as age and sex. These predictions are borne out in our study, in which it can be observed how the group of 12 women in the sample presented higher levels of both anxiety and depression than the men, with the differences in the depression indicators being greater and statistically significant.

The results of this study also indicate that the social support perceived by the total sample was highly positive. Several authors have described how the relationship between social support and terminal illness is associated with extremely difficult and "stressful" situations faced by the patient (symptoms, examinations and prognoses). The data indicate that our patients feel strongly supported by their family and socially, with a mean perceived support score of around 50.39 from a maximum of 55, the mean indicators for the general population being 35.55 (de la Revilla et al., 1991). Obviously, having a sufficient (minimum) level of family support is one of the prerequisites for admission to the home care unit, where the palliative care administered to the patient also requires the attention, time and care of one or more relative(s). It is therefore not surprising that the patients in our sample reported very high levels of social support. If, as the results indicate, the sample of patients report a high and positive degree of social support, this data could be placed alongside the anxiety and depression indicators observed to become another explanatory variable for the acceptance of and adaptation to the illness and its course. Future research should undertake to analyze and confirm this assumption, considering the relationship between the two variables with samples similar to the one used in this study.

A notable and surprising aspect of our results concerns the data on the quality of life variable obtained by means of the QLQ-C30 questionnaire, on the one hand, and the on the various symptoms of the illness as evaluated by the SCS, on the other. According to the EORTC questionnaire, the patients report more problems in relation to symptoms such as *appetite* and *fatigue* than for symptoms such as *constipation* and *pain*. Apparently, the absence of control over these symptoms should have a negative influence on quality of life. However, when patients are asked to assess their *global quality of life*, all of them put it at an intermediate level (neither *terrible* nor *excellent,* i.e., "normal"). On the other hand, the evaluation carried out by the doctor using the SCS revealed that *dyspnea* (what the QLQ-C30 calls *fatigue*) is not precisely the variable that most concerns or incapacitates patients, and that the most incapacitating symptoms would relate *to pain*, *constipation and sleep*, in that order. We can find no explanation for this divergent result, since, despite using different evaluators (the doctor in the checking of symptoms and the psychologist in the QLQ-C30), the patients always answered in the presence of both professionals. In this way we assumed we could discard the presence of different response tendencies depending on whether the evaluation was carried out by a doctor or by another person.

In a similar line, we shall now move on to a discussion of the differences reported by the patients themselves with respect to the same variables after admission to the HCU. In none of the scales of purely psychological content were any differences observed subsequent to the palliative care administered. The initial levels of anxiety, depression and social support are maintained at the second measurement, regardless of the time period between the two evaluations. This is consistent with what was expected, given that in the initial evaluation we observed no maladjustments susceptible to modification. In contrast, all the symptoms reported by the patients to the doctor (pain, constipation, sleep and vomiting) did change. These changes were in a positive direction, with considerable reductions in all symptoms by comparison with the initial levels. Symptoms such as pain, constipation and sleep changed significantly and clinically in the post-evaluation, regardless of when this was carried out. The data and verbal information provided by the patients bear this out.

With regard to the QLQ-C30, no differences were detected between the evaluations in any of the functional subscales, nor in the majority of the symptom scales. It seems that nothing changes after admission to the unit. Neither does the *global quality of life* subscale change with respect to the initial indicators. The variable defined as activity or mobility in the SCS is deserving of special mention. The decrease in activity level of the patients with respect to the first evaluation, despite not being significant ($p<.065$), indicates how the patients in the study experience progressively lower levels of autonomy and mobility. Nevertheless, this is to be expected, and is in keeping with their prognosis, given that we are dealing with a sample of oncology patients in an *advanced* stage of *terminal* illness. In contrast, the *physical functioning* scale, evaluated by the QLQ-C30 and comparable to the *activity* variable of the SCS, given that it also evaluates the area relating to patients' activity and mobility, does not register any change with respect to initial levels. If we bear in mind the terminal prognosis of the patients and the fact that their Karnofsky Index does not improve subsequent to admission in the HCU, it is not unreasonable to think that the "activity" variable evaluated by the doctor is not only more discriminative but also more pertinent than the "physical functioning" scale of the QLQ-C30, at least in a terminal oncology population. In fact, 3 of the 5 items making up this QLQ-C30 subscale (see Appendix 3) are not applicable to a sample with an initial Karnofsky Index of 40, defined as "requires considerable assistance and frequent medical care" (Karnofsky and Burchenal, 1949).

By way of conclusion, it can be said that the term *quality of life* in terminal illness is a vague concept and has been used with different criteria in very different situations (Enck, 1990; Rosenthal, 1993). One of the most widely used questionnaires for evaluating this concept, the QLQ-C30, does not permit the detection of differences subsequent to admission in the HCU in important terminal process symptoms such as vomiting, constipation and patients' quality/quantity of sleep. Nor does it permit the detection of differences observed and reported by patients with respect to the decrease in their levels of activity/mobility. Some of the items included in the scale are not pertinent to the terminal process (e.g., "Do you have any difficulty in going for a long walk?" or "Are you fully capable of doing your job or performing household tasks?"), while some of those that are indeed relevant ("Do you feel depressed?" or "Do you feel nervous?") are too general to provide appropriate information about the patient's level of anxiety or depression. Finally, it is surprising that the patients situate themselves at an "intermediate point" (neither terrible nor excellent) with respect to their *global quality of life*, especially when one of the items of this subscale ("How would you rate your general physical condition during the past week?") should produce a low score, if we accept the veracity of the prognosis and the initial Karnofsky Index. We thus feel that, rather than speaking of quality of life in terminal cancer, it would be more appropriate to speak of degree of comfort achieved and demanded by patients. According to what can be deduced from this study, it is the symptoms of the illness that most concern and incapacitate terminal patients. It seems that they have accepted their illness – which is consistent with data already published in previous studies (Hinton, 1999) and with the levels of anxiety and depression reported in the study –, and that what they are asking for at this point is not an increase in "quality of life", but rather that they be given the assistance necessary to achieve a dignified death without suffering: Quality of death?

"*…For him, all this occurred in an instant and the significance of that instant would never change now. For those present, the death throes continued for two hours more. Something bubbled in his chest and his exhausted body shuddered…*

- It's over! - said one of them, standing over him

He heard these words and repeated them in his soul. "Death is over – he told himself…"

Leon Tolstoy, *The Death of Ivan Illich.*

APPENDIX 1
QLQ-C30

We are interested in knowing some things about you and your health. Please respond personally to all the questions by circling the number that best applies to your case. There are no “correct” or “incorrect” answers. The information you provide shall remain confidential.

	No	Yes
1. Do you have any difficulty in doing activities that require considerable effort, such as carrying a shopping bag or suitcase?	1	2
2. Do you have any difficulty in going for a long walk?	1	2
3. Do you have any difficulty taking a short walk outside?	1	2
4. Do you have to spend the greater part of the day in bed or sitting down?	1	2
5. Do you need help with eating, getting dressed, washing or going to the toilet?	1	2
6. Do you have any problem to do your job or carry out household tasks?	1	2
7. Are you totally incapable of working in a profession or doing household tasks?	1	2

DURING THE LAST WEEK

	Notat all	A little	Quite a lot	Alot
8. Have you had asphyxia?	1	2	3	4
9. Have you had pain?	1	2	3	4
10. Have you had to stop in order to rest?	1	2	3	4
11. Have you had difficulty sleeping?	1	2	3	4
12. Have you felt weak?	1	2	3	4
13. Have you lacked appetite?	1	2	3	4
14. Have you felt nauseous?	1	2	3	4
15. Have you vomited?	1	2	3	4
16. Have you been constipated?	1	2	3	4

DURING THE LAST WEEK

	Notat all	A little	Quite a lot	Alot
17. Have you had diarrhoea?	1	2	3	4
18. Have you felt tired?	1	2	3	4
19. Did any pain interfere in your daily activities?	1	2	3	4
20. Have you had difficulty in concentrating on things like reading the newspaper or watching television?	1	2	3	4
21. Have you felt nervous?	1	2	3	4
22. Have you felt worried?	1	2	3	4
23. Have you felt irritable?	1	2	3	4
24. Have you felt depressed?	1	2	3	4
25. Have you had difficulty in remembering things?	1	2	3	4
26. Has your physical state or medical treatment affected your family life?	1	2	3	4
27. Has your physical state or medical treatment affected your daily activities?	1	2	3	4
28. Has your physical state or medical treatment caused you financial problems?	1	2	3	4

In the following questions, please put a circle around the number 1 – 7 that best applies to you.

29. How would you rate your general physical condition during the last week?	1	2	3	4	5	6	7	Terrible	Excellent
30. How would you rate your general quality of life during the last week?	1	2	3	4	5	6	7	Terrible	Excellent

APPENDIX 2
CORRECTION INSTRUCTIONS

In the symptoms section a number from 0 to 4 should be written according to the intensity of the symptom, bearing in mind the following:

Activity
0: Goes out. Total autonomy
1: Goes out but with assistance
2: Limited to the home but with autonomy in going to the toilet and eating
3: Limited to the bed – chair
4: Bedridden more than 80% of the day

Vomiting
0: No nausea or vomiting
1: Sporadic nausea
2: Vomiting once a day and not every day
3: Vomiting more than once a day
4: Total intolerance

Dyspnea
0: No dyspnea
1: Dyspnea on moderate effort
2: Dyspnea on minimal effort
3: Dyspnea at rest
4: Incapacitating dyspnea

Constipation
0: No
1: Yes

Sleep
0: Well
1: Reasonably
2: Badly

Pain
0: None
1: Slight
2: Moderate
3: Severe
4: Incapacitating

REFERENCES

Aaronson , N.K., Ahmedzai, S.A., Bullinger, M., Cull, A. Duez, N.J., Filiberty, A., Flechtner, H., Fleishman, S.B., de Haes, J.C., Kaasa, S., Klee, M., Osoba, D., Razavi, D., Rofe, P.B., Schraub, S., Sneeuw, K., Sullivan, M. & Takeda, F. (1993). EORTC Core Quality of Life Questionnaire: interim results of an international field study. In D. Osoba (Ed), *The Effect of Cancer on Quality of Life*. Boston: CRC Press, 185-204.

Arrarás, J.I., Illaramendi, J.J. & Valerdi, J.J. (1995). El cuestionario de calidad de vida para cáncer de la EORTC, QLQ-C30. Statistical validation study with a Spanish sample. *Revista de Psicología de la Salud, 7* (1), 13-31.

Bergner, M., Bobbitt, R. & Pollard, W. (1976). The sickness impact profile. *Med Care, 19,* 787-805.

Bredart, A., Didier, F., Robertson, C., Scaffidi, E., Fonzo, D., Costa, A., Goldhirsch, A. & Autier, P. (1999). Psychological distress in cancer patients attending the European Institute of Oncology in Milan. *Oncology, 57* (4), 297-302

Broadhead, W.E., Gehlbach, S. H., De Gruy, F., Kaplan, B.H. (1988). The Duke-UNC functional social support questionnaire. Measurement of social support in Family Medicine patients. *Med Care, 26*, 709.

Carnrike, C.L. & Carey, M.P. (1990). Assessing nausea and vomiting in adult chemotherapy patients: review and recommendations. *Ann Behav Med, 12*, 79-85.

Caro, I. e Ibáñez, E. (1992). La Escala Hospitalaria de Ansiedad y Depresión. *Boletín de Psicología, 36*, 43-69.

De la Revilla, L., Bailón, E., de Dios, J., Delgado, A., Prados, M. & Fleitas, L. (1991). Validación de una escala de apoyo social funcional para su uso en la consulta del médico de familia. *Atención Primaria, 8,* 9, 688-691.

Enck, R.E. (1990). Prognostication of survival in hospice care. *Am J Hosp Palliat Care*, 11-13.

García, G., Cueto J., Arce, C., González, A. (1995). La hospitalización a domicilio: una alternativa a la hospitalización convencional. *Medicina Integral, 25,* 151-155.

Hinton, J. (1999). The progress of awareness and acceptance of dying assessed in cancer patients and their caring relatives. *Palliative Medicine, 13* (1), 19-35.

Hunt, S.M., McEwen J., McKenna, S.P. (1985). Measuring health status. *JR Coll Gen Pract, 35*, 185-188.

Kaasa, S., Bjordal, K., Aaronson, N., Moum, T., Wits, E., Hagen, S., Kvikstad, A. & The EORTC Study Group on Quality of Life. (1995). The EORTC core quality of life questionnaire (QLQC-30): validity and reliability when analyzed with patients treated with palliative radiotherapy. *European Journal of Cancer, 31 A*, 2260-2263.

Karnofsky, D.A. & Burchenal, J.H. (1949). The clinical evaluation of chemotherapeutic agents in cancer. In C.M. Macleod (Ed). *Evaluation of Chemotherapeutic Agents.* New York: Columbia U. Press.

Keizer, M.C., Kozak, J.F. & Scott, J.F. (1992). Primary care providers perceptions of care. *J Palliat Med, 8*, 8-12.

Lougth, M.E., Lindsey, A.M., Shinn, A.J. & Stotts, N.A. (1985). Life satisfaction following heart transplantation. *J Heart Transplant, 4,* 446-449.

Mahoney, F. & Barthel, D. (1965). Functional evaluation: The Barthel index. *Md Med J, 14*, 61-65.

McMillan, S.C. & Mahon, M.(1994). A study of quality of life of hospice patients on admission and at week 3. *Cancer Nurs, 17*, 52-60.

Padierna, C. & Fernández, C. (2001). Instrumentos de evaluación de calidad de vida en pacientes oncológicos terminales: revisión bibliométrica (1988-2000). *Oncología, 24 (5),* 235-246.

Pratheepawanit, N., Salek, M.S. & Finlay, I.J. (1999). The applicability of quality-of-life assessment in palliative care: comparing two quality of life measures. *Palliative Medicine, 13,* 325-334.

Rees, E., Hardy, J., Ling, J., Broadley, K. & A' Hern, R. (1998). The use of the Edmonton Symptom Assessment Scale (ESAS) within a palliative care unit in the U.K. *Palliative Medicine, 12*, 75-82.

Rosenthal, M.A., Gebski, V.J., Keffrird, R.F. & Stuart-Harris, R.C. (1993). Prediction of life-expectancy in hospice patients: identification of novel prognosis factors. *Palliative Medicine, 7,* 199-204

Skarstein, J., Aass, N., Fossa, S., Skovlund, E. & Dahl, A (2000). Anxiety and depression in cancer patients: relation between the Hospital Anxiety and Depression Scale and the European Organization for Research and treatment of Cancer Core Quality of Life Questionnaire. *J Psychosom Res, 49* (1), 27-34.

Snaith, R. P., Baugh, S.J., Clayden, A.D, Hussain, A. & Sipple, M. (1982). The Clinical Anxiety Scale: A modification of the Hamilton Anxiety Scale. *British Journal of Psychiatry, 141*, 518-523.

Spielberger, C.D. (1977). State-trait anxiety and interactional psychology. In D. Magnuson & N.S. Endler (Eds). *Personality at the crossroads: current issues in interactional psychology.* Hillsdale: LEA.

Sprangers, M.A., Cull, A., Bjordal, K., Groenvold, M. & Aaronson, N.K. (1993). The European Organization for Research and Treatment of Cancer approach to quality of life assessment: guidelines for developing questionnaire modules. *Quality of Life Research, 2* (4), 287-295.

Taylor, J.A. (1953). A personality scale of manifest anxiety. *Journal of Abnormal and Social Psychology, 48,* 285-290.

Tierney, R.M., Horton, S.M., Hannan, T.J. & Tierney, W.M. (1999). Relations between symptom relief, quality of life, and satisfaction with hospice care. *Palliative Medicine, 12*, 333-344.

Zigmond, A.S. & Snaith, R.P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatr Scand, 67,* 361-370.

Zung, W.W. (1965). A self-rating depression scale. *Archives of General Psychiatry, 12*, 63-70.

Copyright 2002 by the
Colegio Oficial de Psicólogos. Spain

Psychology in Spain, 2002, Vol. 6. No 1, 77-82

METHODOLOGICAL FORMS AND CUSTOMS IN SPANISH PSYCHOLOGY: AN ANALYSIS THROUGH THE LIFE OF *PSICOTHEMA* (1990-1999)

Ignacio Montero and Orfelio G. León
Universidad Autónoma, Madrid

El presente trabajo explora de forma descriptiva el tipo de metodologías que los investigadores de la Psicología española usan más frecuentemente. Para ello se ha analizado la producción de la revista Psicothema (1990-1999) a modo de ejemplo. Se ha elaborado un sistema de categorización de los diferentes tipos de diseños tomando como unidad de análisis cada estudio, independientemente del número de estudios que contuviera cada artículo. Se ilustran las proporciones de uso de ocho tipos de trabajos: teóricos, experimentales, cuasi experimentales, descriptivos mediante observación, descriptivos mediante encuestas, instrumentales, descriptivos de casos y experimentos de caso único. Dichas proporciones se analizan en función de su evolución a lo largo de la década. Se finaliza discutiendo las implicaciones que, para la enseñanza, pudieran tener la variabilidad metodológica y la presencia de errores en la interpretación de los diseños complejos con interacciones significativas.

The present paper explores in a descriptive way the kind of research methods most frequently used by Spanish researchers in psychology. The production of Psicothema (1990-1999) was analyzed as illustrative of work in this discipline in Spain. A coding system was developed for the categorization of different types of research designs. 'Study' rather than 'paper' was considered as the unit of analysis, given the fact that some papers included several studies. Usage proportions of eight types of studies are illustrated: theoretical, experimental, quasi-experimental, observational descriptions, survey studies, instrumental studies, case studies and one-subject experimental designs. These proportions are presented in relation to year of publication. Finally, we discuss the implications for teaching of issues such as methodological variability and errors in the interpretation of interactions.

Apart of the authors' work being to teach the rudiments of methodology to future psychologists – perhaps future researchers, who knows? – it occurred to them to take a systematic and replicable look at the recent production of researchers in psychology in Spain. Given the boom in publication over the last fifteen years in our country, it has been necessary to use some kind of heuristic bias – perhaps that of accessibility – that would permit us to accomplish a task that would otherwise have been impossible with the available resources. This bias was the coincidence with the ten years of life of the journal *Psicothema*, which, in addition to being well regarded among researchers and having a high degree of impact, has the peculiarity of publishing work in all fields of psychology, which is advantageous for the objective of this study.

This objective, then, is to present a description of the methodologies, the research designs, used in the issues of *Psicothema* over the ten-year period from 1990 to 1999, in the knowledge that, despite being a biased sample, it may be illustrative of the forms and customs referred to in the title of this work (for an analysis dealing more directly with the content of the research published in this journal, the reader should consult Moreno and Sánchez, 1998). Given that our objective is of a descriptive nature, we use the variable time – calendar year – as a basic organizer of this description, which we carry out using the terminology employed in a textbook recently published in our language (León and Montero, 1997; see Kerlinger and Lee, 2000; Shaughnessy, Zechmeister and Zechmeister, 2000, as recent manuals in English).

METHOD

Unit of analysis

The unit of analysis we decided to use for this documentary research was "study", considering as independent units each one of the studies published within a single article. Review articles were considered as theoretical studies. We did not include publications presented within the section "methods, programmes and instruments."

The original Spanish version of this paper has been previously published in *Psicothema*, 2001, Vol. 13. No 4, 671-677

....

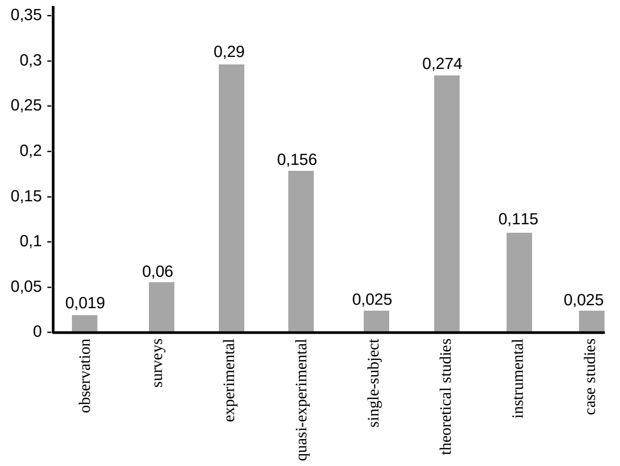
Correspondence concerning this article should be addressed to Ignacio Montero García-Celay. Facultad de Psicología. Universidad Autónoma de Madrid. 28049 Madrid. Spain.
E-mail: nacho.montero@uam.es.

Materials

In order to carry out the descriptive analysis we designed a system of codification, derived largely – as we pointed out earlier – from the terminology of León and Montero (1997). The system was made up of eight main categories, each of which contained several subcategories. The eight main ones, in addition to that of theoretical studies already mentioned, were descriptive studies by means of observation, descriptive studies by means of surveys, experimental studies, quasi-experimental studies, single-subject experiments, instrumental studies (development of tests and equipment, design and/or adaptation of these) and descriptive case studies. Appendix A includes a definition of each category and its subcategories.

Reliability of the categorization system was studied by calculating the percentage of agreement between coders. An advanced psychology student was trained to use the coding system and as a sample we took, at random, all the studies published in a particular year. With these studies categorized independently by the student and one of the authors, an agreement level of 95% was obtained. In cases of disagreement, the inclusion criteria were clarified until agreement was reached. On classifying the publications from another year, an agreement of 99% was achieved.

Figure 1
Summary of the methods used between 1990 and 1999



Design and procedure

This work is a descriptive study by means of the analysis of documents. The description was carried out using a system for classifying the methodology employed in each of the studies analyzed. Once this system had been developed, its reliability was tested using the procedure referred to in the previous section. Furthermore, we used the variable time as the axis of the description, the calendar year of publication of each study constituting the form of categorizing this variable.

RESULTS AND DISCUSSION

Table 1 shows the distribution of the number of studies considered for our research, grouped according to year of publication. The marked increase from 1993 to 1994 is due to the fact that the journal went from publishing two numbers per year to publishing three. We should like to stress that the studies in the section “methods, programmes and instruments” were not taken into account. Total number of articles published in the decade studied was 367. This indicated an annual mean of almost 37, if we take the decade as a whole. If we separate the publications in two periods according to the number of issues per year, we find that the first period (1990-93) has an annual mean of 23 articles published, 11.5 per number, while the second period (1994-99) shows a mean of almost 45.8 studies each year, a mean of over 15.3 per issue. This implies an increase in the number of studies published that is independent of the increase in number of issues: the mean of 23 articles with two numbers almost doubles to reach 45.8 with three issues per year.

With regard to the type of methodology used in each one of the studies, Figure 1 shows a histogram of the distribution across the decade. The proportion of reviews (theoretical studies) is 0.274, more than a quarter of the total. Almost another third of the publications is accounted for by experimental studies (0.29), though it must be borne in mind that this group includes studies that analyze variables not always manipulated by the researcher. It was sufficient for one variable to be manipulated for the study to be included in this category. The next group of studies as regards frequency is that which includes those labelled quasi-experimental, which account for 0.156 of the total, a little over half the share represented by the previous group. This group is followed by those studies

Table 1
Distribution of units of analysis by year of publication

Year	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	Total
Studies	18	22	30	22	33	41	47	49	54	51	367

we call instrumental, and which involve the development of tests and equipment and their design and/or adaptation.The proportion in this case is 0.115. Descriptive studies by means of surveys account for 0.06 of the total, half as much as the instrumental studies and less than a quarter of the proportion represented by experimental studies. Single-subject experimental studies and case studies each represent proportions of 0.025, while observational descriptions occupy the smallest proportion, at 0.019. We find, therefore, that approximately 42% of the publications with empirical content are experimental in nature, and almost 17% are instrumental studies. Between them, these two types of study account for almost two-thirds of the articles with empirical content.

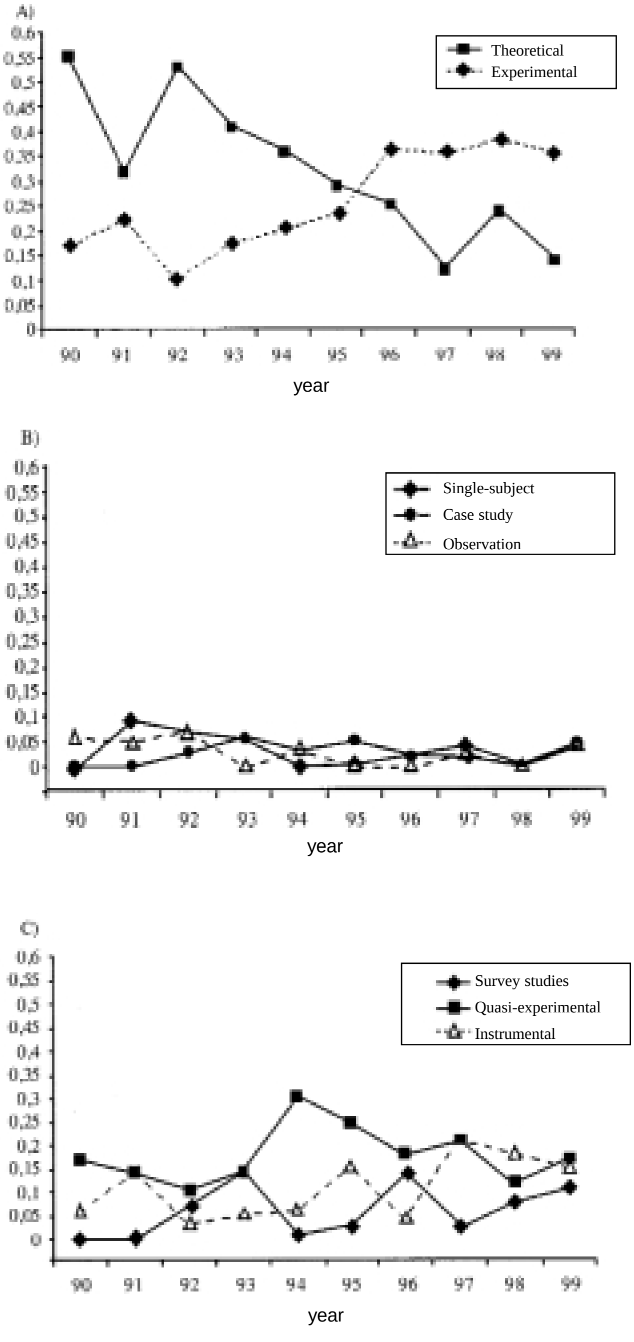
In the following figure we try to illustrate the evolution of the different types of study. Figure 2a shows the changes in the number of theoretical studies in relation to the experimental articles. The theoretical works initially account for a very high proportion – 0.55 in 1990 –, but this proportion falls to between 0.10 and 0.15 in the last two years studied. In contrast, the experimental studies are poorly represented in the first three years, with 0.10 in 1992, but consolidate themselves at a proportion of 0.35 or over from 1996. The evolutions of these two types of study cross around 1995, at which point they each account for a proportion of around 0.25.

Figure 2b shows the evolution of the observational descriptions, the descriptive case studies and the single-subject experiments. These three types of study have in common that they account for low proportions – always less than 0.1 – and show a certain stability throughout the decade.

The quasi-experimental studies, the survey studies and the instrumental studies (see Figure 2c), on the other hand, present a more oscillating profile over the decade. The quasi-experimental works move within a range from 0.1 to 0.3, the survey studies between zero and a proportion close to 0.15 (though their annual mean is 0.06), and the instrumental studies between a minimum proportion of 0.03 and a maximum of 0.20, with a clear difference in favour of the second part of the decade.

Having described the main categories of the coding system employed, we shall now present some data showing how these broad categories are broken down into more detailed ones, bearing in mind that this breakdown can help to better illustrate the comments made up to now. This breakdown affects the experimental and quasi-experimental studies. As far as the first group is concerned, Figure 3 shows the separate evolution of studies with simple experimental designs – of a single independent variable – and factorial designs – with more than one independent variable, at least one being manipulated. As a general comment, we should point out that the crossing of the tendencies in the final year of the analysis may be insignificant, though to be sure of this we would need to continue the data series further. This finding aside, it would appear that, as is to be expected,

Figure 2 a:
Comparative evolution of the frequency of theoretical and empirical studies; b: Comparative evolution of the frequency of single-subject experiments, case studies and observation studies; c: Comparative evolution of the frequency of survey, quasi-experimental and instrumental studies



there are more studies with more than one independent variable. The relationship between one type of study and the other – ignoring the years in which no simple experiments were published –, goes from parity in 1993 to a ratio of four to one in 1994 and three to one in 1998. We should add that the mean number of independent variables included in the factorial studies (the experimental ones plus the complex prospective “ex post facto” ones) is 2.56, the vast majority being of two (45 studies) or three (26 studies). Eight studies with four independent variables were published, and one with five.

A separate mention should be made of another fact that is not reflected in the figure, but is nevertheless relevant. We are referring to the study of interactions, and the way in which this affects the interpretation of the lower-order effects, be they simple interactions – when there are at least three independent variables – or main effects. Of the 80 studies analyzed, 46 report the presence of some significant interaction. In 27 of them (58.7%), the interpretability of the lower-order effects may be affected. Lack of sufficient information means that we cannot say for certain that it is affected in all cases, since a significant interaction does not always lead to error in the interpretation of lower-order effects. However, in 15 of these studies we can affirm that the interpretations made of the lower-order effects are incorrect. The errors are of two types. The first type consists in interpreting simple interactions directly once the presence of a higher-order interaction has been detected. To do so is erroneous, because what this interaction means is that at least one of the possible simple interactions does not occur in the same way for all the levels of the variable that is withdrawn from the analysis on passing from one order of interaction to the one immediately below it. The other type of error is simpler to explain. Depending on the type of double interaction produced, the main effects may constitute an inadequate summary of the simple effects. Clearing this up involves studying the simple effects. The cases we mention are those where the effect of an independent variable is reported through the interpretation of its main effect when this is not an adequate summary. This occurs when the influence of the independent variable on the dependent one differs as a function of the levels of the second variable involved in the significant interaction, that is, when qualitatively different simple effects occur (see León and Montero, 2001).

Figure 3
Types of experimental designs, by number of independent variables

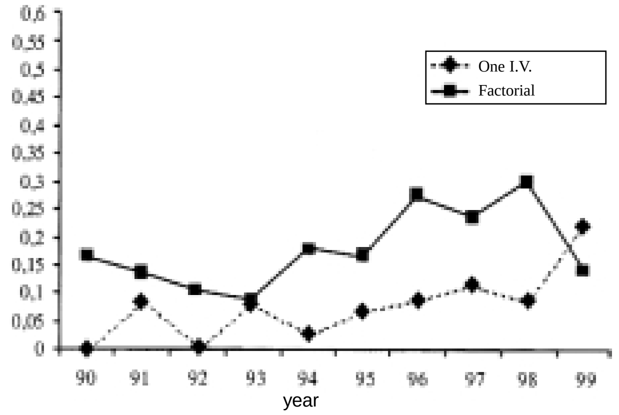
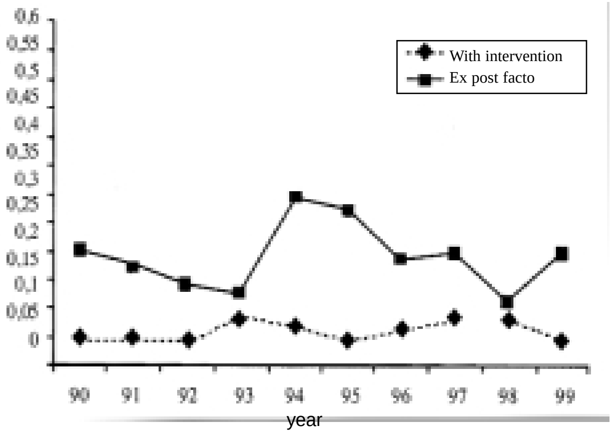


Figure 4
Types of quasi-experimental designs, by type of intervention



As regards the articles that include studies of a quasi-experimental nature, Figure 4 breaks them down into two main types: designs with intervention and “ex post facto” designs. With a proportion that hovers around 0.15, “ex post facto” designs are always more common than designs with intervention, which remain around a proportion figure of 0.05.

Figure 5 shows that, within the “ex post facto” group, prospective designs (0.72) are more common than retrospective ones (0.25). Of the prospective studies, the most frequently used type is simple (0.33). León and Montero (1997) define a simple prospective design as that which studies a single independent variable that cannot be manipulated by the researcher within a research context in which first of all groups are selected according to their value in the variable under study and subsequently their differences in the dependent variable are studied. In the case of the retrospective studies, the “single group” type is the most numerous (0.14). The manual we have just cited defines a single-subject retrospective design as that carried out in a context in which, within a sample as large and representative as possible, researchers measure simultaneously the dependent variable and all those variables that are candidates for explaining, with a minimum of significance, its variability. In this last figure the proportions relate to the total of “ex post facto” type studies.

Another relevant point that arose as a result of our analysis concerns the fact that we found excessive variation in the forms of presentation of the studies, which is surprising, since the journal requires authors to follow the APA norms. As this issue is closely related to the task of reviewers, a series of recommendations for them is presented in Appendix B.

CONCLUSIONS

Although the nature of this work is purely descriptive and the analysis of a single journal precludes any temptation to generalize, we feel it is possible to make some remarks and raise a few questions by way of conclusion.

First of all, we are concerned about the degree of methodological variability found. We ask ourselves whether it is not somewhat scarce, given the quantity of research methods that are considered as fundamental subjects in psychology teaching curricula. Let us recall that two-thirds of the studies with empirical content were classified within the categories of experimental and instrumental studies. And while we are alarmed about this bias as reflected in the journal, we are even more concerned about its practical implications for our teaching. We feel that the justification of the importance of our disciplines in the training of future psychologists (given that they will become, if not researchers, at least consumers of quality research) is somewhat diminished, if we are to judge by the content of the journal in question. And with regard to the training of future researchers, we are tempted to make the following observation: in optional subjects and doctoral programmes it may not be necessary to explain important methodological developments, in view of the use made of them by those who actually conduct psychological research in our country. Clearly, it is positive to always widen and sharpen students’ knowledge. Frankly, however, the apparent distance between what we teach and what is used amply covers this anticipatory function, at least, we insist, in the light of what we have found in this decade of *Psicothema*.

In addition to the above, another observation we consider to be important concerns the interpretation of lower-order effects when significant interaction effects are found. Although there is a long tradition in the study of errors in data analysis (see Judd, McClelland and Culhane, 1995; Keselman et al., 1998), what we wish to focus on here is related not to the way we *decide* about hypotheses, but to the way we *interpret* the results. At present there is still no clear explanation of the origin of these potential errors of interpretation, which, moreover, can be observed in any Western psychological research journal (see León and Montero, 2001). What is a fact is that in the analysis presented here, the percentage of cases in which they may appear is high (around 60%). The fact that in a third of all the studies we have found not only the possibility of error occurring but its actual occurrence reflects the relevance of the problem. Although awareness of the problem may be a starting point, an “insight” is not necessarily intrinsically therapeutic.

And it is in this way that the two aspects discussed in these conclusions are related. It would seem necessary to call for reflection on the scope of what we teach in our methodological disciplines since, in the light of the present analysis, few different research approaches are used and, on some occasions, their results are interpreted quite incorrectly. One proposal might be the following: “Let us teach less content, but let us teach it better.”

ACKNOWLEDGEMENTS

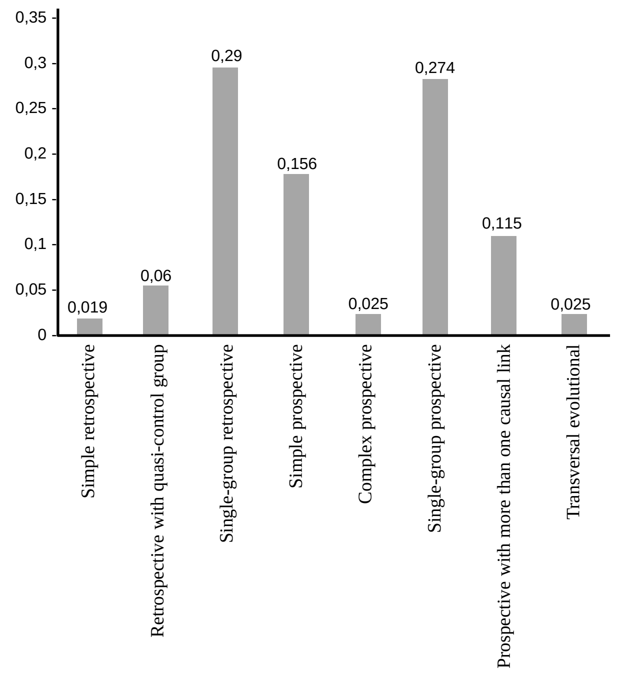
The authors of this work would like to thank Beatriz Plans for her excellent work in the classification of the studies analyzed.

REFERENCES

A.P.A. (1994). *Publication manual of the American Psychological Association* (4th ed.) Washington, DC: APA.

Judd, C.M., McClelland, G.H. & Culhane, S.E. (1995). Data analysis: Continuing issues in the everyday

Figure 5
Types of “ex post facto” designs



analysis of psychological data. *Annual Review of Psychology, 46,* 433-465.

Keselman, H.J., Huberty, C.J., Lix, L.M., Olejnik, S., Cribbie, R.A., Donohue, B., Kowalchuk, R.K., Lowman, L.L., Petosky, M.D., Keselman, J.C. & Levin, J.R. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA and ANCOVA analyses. *Review of Educational Research, 68,* 350-386.

Kerlinger, F.N. & Lee, H.B. (2000). *Foundations of behavioral research* (4th ed.). Fort Worth, TX: Harcourt College Publishers.

León, O.G. & Montero, I. (1997). *Diseño de investigaciones. Introducción a la lógica de la investigación en Psicología y Educación* (2ª ed.). Madrid: McGraw-Hill Interamericana de España.

León, O.G. & Montero, I. (2001). Cómo explicar el concepto de interacción sin estadística: Análisis gráfico de todos los casos posibles en un diseño 2x2. *Psicothema, 13.* 165-171.

Moreno, S. & Sánchez, A. (1998). Análisis bibliométrico de la revista “Psicothema” (1989-1997). *Psicothema, 10.* 23-27.

Shaughnessy, J.J., Zeichmeister, E.B. & Zeichmeister, J.S. (2000). *Research methods in Psychology* (5th ed.). New York: McGraw-Hill.

APPENDIX A

CATEGORIES USED IN THE CLASSIFICATION SYSTEM

The system of classification of the methodologies used in the different studies analyzed was developed on the basis of the terminology proposed by León and Montero (1997). Below we present the eight main classification categories and their subcategories, where applicable. Given that the denominations are sufficiently well known, we only provide inclusion criteria in cases where, in the literature, the denominations are not common from author to author.

Theoretical study: we included in this category all the articles that did not provide empirical data generated by the authors. That is, all reviews that did not report an empirical study.

Observational descriptions: This category includes studies that used natural or structured systematic observation, with a descriptive objective.

Survey study: This group included all the studies that used surveys with a descriptive purpose, indicating the type of design used, transversal or longitudinal.

Case study: We considered as case studies all those dealing with a single sample unit, be it a person, group, organization or other, and whose objective was descriptive.

Experimental study: In order for a study to be classified as experimental it was necessary for at least one of the variables studied as independent to have been manipulated by the researcher. Experiments were categorized as simple – a single independent variable – or complex – more than one independent variable. We also noted the nature of each one of the independent variables – manipulated or not manipulated – and the type of design – between- or within-subjects – employed with each one.

Quasi-experimental study: In this category we included all those studies that, despite having the objective of checking a hypothesis of causal relation, had limitations – more or less serious – for successfully accomplishing that objective. We included here designs with intervention – applications in natural situations in which it is impossible to assign subjects at random or control the order of application of the levels of the independent variable – and “ex post facto” studies, characterized by the impossibility of manipulating the independent variable. Within the designs with intervention we took into account each of the thirteen variants listed in León and Montero (1997). We made the same consideration for the “ex post facto” studies.

Single-subject experimental study: In this category we included all the studies that used existing experimental techniques for application to single cases. Five variants were found: AB, ABAB, multiple baseline for various behaviours of the same subject, multiple baseline for the same behaviour in various situations, and multiple baseline in various cases.

Instrumental study: We considered as belonging to this category all the studies whose purpose was to develop tests and equipment, including both their design and adaptation.

APPENDIX B

REFLECTIONS FOR REVIEWERS

These lines are addressed to reviewers of the journal and not to authors, since it is the former that decide whether a manuscript is ready for publication. We feel that the communicability and replicability of reports would be improved if reviewers were to take into account the following and to inform the authors in a detailed and didactic way. We should stress that the comments that follow refer to exceptions, and do not correspond to either the average or the mode of the publications.

It is obvious that the design serves to achieve an objective made explicit in the introduction – so obvious that we as authors often think it unnecessary to explain it or justify it. The reader tends to have a different point of view.

Common ground: we do not like norms. This conclusion would be reached by anyone that reads the set of “methods” sections from these ten years of journals. If those who write and review studies were to follow the APA norms (1994) required by the publication, this paragraph would be unnecessary. Specifically, we feel there would be an improvement if we insisted upon: a) naming the design that is to be used; stating whether the variables will be between- or within- and indicating the corresponding number of levels and their names; b) making explicit how groups were formed – making it clear to authors that not using a biased criterion is not the same as a random assignment; c) avoiding the word “influence” in titles and abstracts when the research is not experimental; d) using the recommended sub-sections in “methods” and not adding others (e.g., statistical procedures); e) dispensing with creativity when making tables: it is easier for the reader to follow a standard format; f) reporting the means and variances of the groups; and h) although it may seem punctilious, writing up the statistical results in a standard way: a different format from each author in the same issue of a journal does not give an elegant result.

Psychology in Spain, 2002, Vol. 6. No 1, 83-95

BODY-IMAGE DISTORTION IN EATING DISORDERS: A META-ANALYSIS

Ana Rosa Sepúlveda, Juan Botella and José Antonio León
Psychology Faculty. Autónoma University of Madrid

Se presentan los resultados de un estudio meta-analítico sobre la naturaleza de la alteración de la imagen corporal asociada a los trastornos de la alimentación. Se admitieron 83 estudios primarios independientes, publicados entre 1970 y 1998, que cumplían con los criterios de selección establecidos, y que dieron lugar a 238 estimaciones de tamaños del efecto. El grado de distorsión estimada en la imagen, mediante el índice d, fue de 0.545 en anorexia, 1.019 en bulimia y 1.185 en bulimarexia. Sin embargo, los resultados revelan que hay una gran heterogeneidad entre los estudios. El factor más frecuentemente asociado es la forma de evaluar la alteración. Las medidas actitudinales, que tienen componentes cognitivo-afectivos, muestran valores superiores a las de las medidas perceptivas. Se analizó la influencia de otras variables moderadoras, como el tipo de instrucción, el estatus del paciente, el instrumento de diagnóstico, etc. Se discuten las posibles explicaciones de los resultados y sus implicaciones clínicas

The results of a meta-analytic study on the nature of body-image distortion associated with eating disorders are reported. A total of 83 independent primary studies were admitted, all published between 1970 and 1998, and fulfilling the previously-established selection criteria. The studies produced 238 estimations of effect size, measured as d index. Degree of distortion of the image was 0.545 for anorexia, 1.019 for bulimia and 1.185 for bulimarexia. However, the analysis revealed great heterogeneity across studies. The factor most frequently associated is the way the distortion is assessed. The attitude measures, which include cognitive-affective components, show greater average values than the perceptual measures. Also analyzed was the influence of other moderating variables, such as type of instructions, patient status or diagnosis procedure. Alternative explanations of the findings and clinical implications are discussed.

Body-image distortion has been identified as one of the diagnostic criteria of anorexia nervosa in eating disorders (American Psychiatric Association, 1994). Given its important role in the onset and maintenance of illness, recovery from it is crucial to treatment (Bruch, 1962; Kolb, 1975). However, one of the problems with regard to this diagnostic criterion is that it has spread to other sectors of the population, giving rise to a marked increase, in general, in preoccupation with physical appearance and/or weight.

Although body image and its distortion in eating disorders has been the subject of a large number of studies, the results have not been consistent, so that there is still no consensus on the nature of the problem. Some of this inconsistency may be due to the use of different diagnostic criteria or different methods of assessment, among other factors. Authors such as Smeets, Smit, Panhuysen and Ingleby (1997) analyzed by means of meta-analysis the influence of different assessment methods in the estimation of body size in anorexic patients. They concluded that Slade's whole-body and visual procedure of size estimation methods assessed aspects of body image that were inter-related. In a later meta-analytic study, these same authors (Smeets, Panhuysen & Ingleby, 1998) studied the relationship between actual body size of anorexic patients and their Body Perception Index (BPI), obtaining a negative linear relationship; that is, the smaller their body size, the greater their overestimation with regard to their body.

By body-image distortion we understand the presence of value judgements about one's body that do not coincide with the true characteristics. Although a certain margin of error is always to be expected in appreciations of one's own body, the presence of systematic biases in patients with eating disorders has led to the generalization of the concept of *body-image distortion*.

Body image is made up of perceptual, cognitive-affective and behavioural components. A review of the empirical literature reveals the two main ways in which it has been attempted to assess these different components, to

---

The original Spanish version of this paper has been previously published in *Psicothema*, 2001, Vol. 13, No 1, 7-16

....

Correspondence concerning this article should be addressed to Juan Botella. Facultad de Psicología. Universidad Autónoma de Madrid. Cantoblanco s/n. 28049 Madrid. Spain. E-mail: juan.botella@uam.es which end a variety of techniques have been created. These two ways are:

(a) Accuracy in the estimation of subject's body size, based on purely perceptual judgements. There are two main forms of assessing it, according to the object of estimation. On the one hand, in some techniques the width of certain parts of the body is measured, such as the face, the hips or the waist. With these data a body image index can be obtained (BPI, *Body Perception Index*) as proposed by Slade & Russell (1973), which relates size estimated by the subject to real size measured by an anthropometer [BPI=(perceived size/real size) x100]. Other frequently used techniques are the *Movable Caliper* (Slade & Russell, 1973) or the Marked Image (Askevold, 1975), as well as whole-body techniques of estimation, such as that of Video Camera (Allebeck, Hallberg & Espmark, 1976) or Silhouettes (Williamson, Kelly, Davis, Ruggerio & Blouin, 1985), to which Slade's index is also applied.

(b) The individual's attitude and feelings towards his or her own body, which reflects attitudinal, affective and cognitive variables. This approach has been developed by means of questionnaires designed to measure attitude to weight and body form, as well as attitude to food, binge-eating and diets; they provide a Body Dissatisfaction Index. There are also more specific questionnaires on body image, such as the BSQ (Cooper, Taylor, Cooper & Fairburn, 1987) and other more general ones, such as the EAT (Garner & Garfinkel, 1979).

The importance of being able to offer a clearer and more consensus-based view of this concept in the field of eating disorders led us to make it the main objective of a meta-analysis. Although it has been attempted to integrate the results obtained in this research field by means of narrative reviews (Steinhausen & Glanville, 1983; Cash, 1995), it is important to complement them with procedures that allow a quantitative overview of the results, as well as to detect and clarify the inconsistencies observed in the primary studies (Glass, 1976; Smith & Glass, 1977; Sánchez & Ato, 1989).

The general objective of the present meta-analysis is to update and complement the results of the meta-analysis by Cash & Deagle (1997) on body image in eating behaviour disorders. Specifically, our meta-analysis not only covers a larger number of studies (83, published between 1970 and 1998, as against the 66 of Cash & Deagle, published between 1974 and 1993), but it also includes more works carried out with European and Asian samples (Cash and Deagle restricted themselves almost exclusively to North-American samples). Moreover, we studied a larger quantity of moderating variables, such as type of instruction, status of patients, type of diagnosis or type of control group used.

OBJECTIVES AND HYPOTHESES OF THE META-ANALYSIS

Our specific objectives are: (A) To integrate in a quantitative manner the results of studies that assess body-image distortion in eating disorders; (B) To analyze a range of characteristics of the studies, as potential moderators of the results; (C) To study the nature and scope of the concept of body image, taking into account the type of component that is dysfunctional: the perceptual component (distortion of body size) and the cognitive component (body dissatisfaction); (D) To suggest therapeutic lines based on the conclusions drawn.

In accordance with the objectives set and the review of the existing literature, we proposed a series of hypotheses referring both to the nature of the body image and to the moderating variables that may have influenced the results. As regards the *nature of the body image*: (A) Following the conclusions and suggestions of various authors (Huon & Brown, 1986; Gleghorn, Penner, Powers & Shulman, 1987; Mizes, 1992; Probst, Vandereycken, Van Coppenolle & Vanderlinden, 1995; Lovell & Williamson, 1997), we expect the cognitive component, assessed as body dissatisfaction, to show a greater effect size than the perceptual component, measured as body distortion; (B) In accordance with previous research (Lindholm and Wilson, 1988; McKenzie, Williamson and Cubic, 1993), we expect to observe greater dissatisfaction with the image and greater perceptual distortion in bulimia than in anorexia, given that in bulimia there is greater discrepancy between true body size and ideal body size. As far as *subject characteristics* are concerned, we propose that the magnitude of the effects will be associated with age and relative body mass index (BMI); in sum, we expect effect size to increase with age and BMI (Halmi, Goldberg & Cunningham, 1977; Birchnell, Lacey & Harte, 1985; Sánchez-Carracedo & Saldaña, 1998).

With regard to the *methodological* variables, we expect to obtain results similar to those of previous studies, in which it was found that body perception techniques show a greater effect size for distortion than techniques based on perception of the body by individual parts (Mizes, 1992; Probst, Vandereycken, Van Coppenolle & Pieters, 1995; Probst, Vandereycken, Van Coppenolle & Vanderlinden, 1995).

We shall also analyze other variables that have aroused theoretical interest more recently, such as type of diagnosis, type of instructions and body part rated as largest in the perceptual assessment, given the possibility that these factors may covary with the effect sizes.

METHOD

Search for information in the literature

The information search was based on the following sources: a) Computerized databases from 1974 to 1998 (PsycLIT, MEDLINE and CSIC), searched in October 1998; b) Direct review of specialized journals, such as *The International Journal of Eating Disorders* up to May 1999 and its references, books and monographs. We analyzed a total of 258 articles, from which we selected 83 studies that fulfilled the following criteria: (A) It included at least one clinical group with known diagnostic criterion of anorexia or bulimia nervosa. We also decided to include a third clinical group of patients with bulimarexia. Although this category is not yet recognized specifically within a diagnostic criterion, it is currently accepted and employed by the medical community to discriminate those patients that have passed through a previous anorexia and developed bulimia, alternating in different periods the symptoms of one or the other; (B) It included one or more control groups. Some studies included two groups of normal controls that differed in that one presented restrained eating behaviour and the other did not; (C) It assessed the presence of body image distortion in cognitive or perceptual aspects or both; (D) It included sufficient information to be able to calculate the effect size; (E) Although it may have included standardized or projective assessments, perceptual aspects should be kept independent of cognitive aspects.

Coding of categorical characteristics and variables

The characteristics of the studies were coded according to nature of the body image and of the dependent variables. The characteristics were distributed in the following categories (Sánchez & Ato, 1989). As *substantive variable*, the aspect of body image studied, categorized as perceptual aspects versus cognitive aspects. The *subject variables* were coded as: (a) type of group by eating disorder (anorexia, bulimia and bulimarexia); (b) age of subjects (patients and controls equal, patients older or patients younger); (c) body mass index (equal, patients higher, patients lower); (d) patient's status (inpatients, outpatients or mixture); (e) status of control group (psychiatric, non-psychiatric). We coded two *contextual variables*: (a) decade of publication of the article (1970-79, 1980-89, 1990-99); (b) origin of the samples (North America, Australia, Europe and Asia). The *methodological variables* coded were: (a) method of body image evaluation (perceptual methods could be whole body or parts of body, and attitudinal questionnaires on dissatisfaction could refer to weight/shape or global attitude); (b) within perceptual methods it was analyzed which part of the body was rated as largest (head, shoulders, hips, waist, stomach or thighs); (c) diagnostic procedure used (Russell, Feighner, DSM-III, DSM-III-R, DSM-IV or various); (d) type of instructions given to subjects in the perceptual evaluation, which was divided into 6 possible categories: *Ambiguous instructions* ("*assess four body parts*" or "*give four trial estimates each of actual frontal and actual profile size*"); *cognitive instructions* ("*how she thought it reflected her real image*" or "*how I think I look*"); *affective instructions* ("*what she felt her body was like*" or "*how I feel I look*"); *both affective and cognitive* (the study specifically mentions results for cognitive and affective instructions); *not specified* (no information); *unnecessary* (studies that use attitudinal methods by means of questionnaire).

Effect size was calculated for each assessment method. With regard to techniques for assessing body parts, we opted to code the score for greatest distortion for a part of the body. This perceptually most distorted part was coded, in turn, as dependent variable. In studies that had obtained results of various measures of body image, an effect size was calculated for each measure separately. In the process of coding of the results it was necessary to adopt different types of judgements, which were resolved by consensus between the two coders.

Measurement of effect size and analysis techniques

With the aim of comparing the results of the studies, we obtained the initial effect size (*d*), before the treatment. For this purpose we calculated the difference between the mean values in distortion in the clinical and control groups, divided by the standard deviation of the control group corrected to avoid estimation bias (Hedges & Olkin, 1985). Effect size reflects whether there is body-image distortion in patients with eating disorders or not. Given that it is the standardized mean difference between patients and controls with respect to a measure, this was calculated on the values before subjects underwent any kind of treatment. If the value of *d* is positive, it means that the group with eating disorders presents greater distortion or dissatisfaction than the normal subjects. However, we should make clear that this says nothing about the absolute value of the distortion – only about the "relative" distortion between the clinical and normal groups.

The effect sizes were calculated from the means and standard deviations. When these statistics could not be obtained directly from the studies, they were obtained manually from other statistics in the article, such as *t*-tests or *F* ratios.

The meta-analysis was carried out by applying the statistical method of Hedges & Olkin (1985), using the Johnson DSTAT (1993). After analyzing in a descriptive way the characteristics of the studies, we took weighted averages (by means of the sample sizes) of the estimations of effect size, and confidence intervals were calculated with 1-a=0.95. Subsequently, we studied the homogeneity of the effect sizes with the Q statistic and we carried out analyses of the moderating variables, by means of analyses of variance, creating the appropriate categorical models, and weighted regression analyses.

RESULTS

Analysis of the characteristics of the participating studies

Table 1 shows a quantitative description of the variables coded in the studies. Several aspects of the data are worthy of particular mention. On the one hand, the similar degree of presence of the clinical groups of anorexia and bulimia, while the bulimarexia group represents just 6%. Sixteen percent of the studies do not specify the diagnostic criteria used for selecting their patient samples, whilst a similar percentage use a variety of diagnostic criteria. The type of instructions most frequently used are cognitive, with 40%, whilst 17% do not specify the type of instructions given to participants. The body parts most commonly used for the evaluations are the face and the waist. With regard to age and BMI, 19% do

Table 1
Description of the moderating variables

Variables	Categories	Frequency	%
Year of publication	1. 1970-1979	9	10,8
	2. 1980-1989	44	53
	3. 1990-1999	30	36.1
Origin of the samples	1. North America	46	55.4
	2. Australia	9	10.8
	3. Europe and rest	27	32.5
	4. Asia	1	1.2
Status of patient	1. Internal patients	25	23.1
	2. External patients	66	61.1
	3. Mixture	9	8.3
	4. Not specified	8	7.4
Clinical group	1. Anorexia nervosa	50	44.6
	2. Bulimia nervosa	55	49.1
	3. Bulimarexia	7	6.25
Status of controls	1. Psychiatric	3	3.3
	2. Non-psychiatric	88	95.6
	3. Not specified	1	1
Diagnosis	1. Russell	6	7.2
	2. Feighner	11	13.2
	3. DSM-III	13	15.6
	4. DSM-III-R	27	32.5
	5. DSM-IV	1	1.2
	6. Various	12	14.5
	7. Not specified/not known	13	15.7
Type of instructions	1. Ambiguous	12	23
	2. Cognitive	21	40
	3. Affective	5	9
	4. Both	6	11
	5. Not specified	9	17
Body part rated largest	1. Face	11	39
	2. Shoulders	1	3.5
	3. Waist	8	28
	4. Hips	3	10.7
	5. Stomach	3	10.7
	6. Thighs	2	7
Body-image measures	1. Perceptual whole-body[1]	30	22.7
	2. Perceptual body parts[2]	36	27.3
	3. Attitudinal weight/shape[3]	32	24.2
	4. Attitudinal global[4]	34	25.7
Age[5]	1. Equal	54	46.1
	2. Patients older	34	29
	3. Patients younger	7	5.9
	4. Not specified	22	18.8
BMI[5]	1. Equal	13	11.1
	2. Patients higher	5	4.2
	3. Patients lower	26	22.2
	4. Not specified	73	62.4

[1] Techniques such as BIDOD, BMS, BID or DPT are coded as global perceptual measures.

[2] Techniques such as BIA, IM, MCT or KSEA are coded as body parts perceptual measures.

[3] Assessment techniques such as EDE, BATH, BMQ, CBT, SRBI, BCS or BISE are coded as weight/shape attitudinal measures.

[4] Assessment techniques such as EDI (DT, D), MAC, RS, EAT, BULIT, BES, SCS, BCDS, BIAQ, BPSQ, BDQ, BSC, BAT, BSS or Stroop are coded as global attitudinal measures.

[5] Age and BMI are considered equal when the mean values are identical or differ by a maximum of 0.5.

not specify the former and 62% fail to specify the latter.

Half of the 238 effects were obtained with questionnaires that assessed body image in a global way (Table 2), with attitudinal questionnaires on body shape and weight representing just 10.1%. In only 20 studies was a single effect size obtained. The mean of effect sizes per study is three, and in three studies as many as 12 effects were obtained, due to the use of several clinical groups and several body-image measures. Mean age of the different clinical and control groups does not differ significantly, the mean being 22 years. As far as BMI is concerned, the lowest mean is that of the anorexia sample (16.3), and the highest mean is that of the control sample (21).

It was considered useful, when there were two non-psychiatric control groups that differed only in level of concern about diet or of restraint in eating, to compare them both with the clinical group.

Summary of estimations of effect size

The effect size for the anorexia group was positive; this means that the anorexia patients distort or feel dissatisfaction to a greater extent than the controls, the weighted mean of the effect size being 0.545. Effect size for the bulimia group was also positive, that is, they also distorted more or were more dissatisfied than the controls, 1.019 being the estimated value. It should be underlined that this group presents higher levels of distortion than that of anorexia. The bulimarexia group obtained the highest value, with 1.185. Table 3 shows the results of the homogeneity tests for the groups, all three of which are significant. In sum, body-image distortion is greater in bulimia and bulimarexia patients, but since the homogeneity tests are significant, we deduce that there is still greater heterogeneity than that which would be expected as mere random fluctuations. We shall therefore move on to studying the possible effects of those moderating variables that may have some influence on these results.

Perceptual measures versus attitudinal measures

With the aim of studying the nature of body image distortion, Table 4 shows the indices for the perceptual estimation of body size (body distortion index) and for those measured by the questionnaire (body dissatisfaction score).

Patients with eating disorders present greater body-image distortion if they employ attitudinal or cognitive measures (*d* = .92) than if they employ perceptual measures (*d* = .55) ($Q_B(1) = 100.5$, $p < .0001$). On carrying out this analysis for each group we find that the anorexia group presents a significantly higher average in the attitudinal measures than in the perceptual ones [$Q_B(1) = 25.739$; $p < .007$]; the same occurred with the bulimia group [$Q_B(1) = 69.710$; $p < .001$]. With regard to the bulimarexia group, the difference is not significant [$Q_B(1) = 1.104$; $p < .293$], but it should be borne in mind that the analysis of this categorical model is based on a considerably smaller number of estimations.

As regards the model for type of clinical group, the anorexia groups differ less then the control groups on evaluating their distortion or body dissatisfaction, while the bulimia groups present a significantly higher level with respect to their controls in distortion or body dissatisfaction [*d* = .53 and *d* = .94; $Q_B(1) = 121.5$, $p < .0001$].

On comparing the anorexia and bulimia groups we obtain a significant effect of type of diagnosis [*d* = .38 and *d* = .70; $Q_B(1) = 29.17$, $p < .0001$], while in the meta-analysis by Cash & Deagle (1997) no significant effect is obtained. Given that the values are ordered in

Table 2
Description of age, BMI and body-image measures

VARIABLE	CATEGORY	
		NUMBER OF EFFECTS
Body-image measures	Perceptual whole body	42 effects (17.6%)
	Perceptual body parts	53 effects (22.3%)
	Attitudinal weight/shape	24 effects (10.1%)
	Attitudinal global	119 effects (50%)
		AVERAGES
Age	Anorexia	21.1
	Bulimia	22.7
	Bulimarexia	22.9
	Controls	21.3
BMI	Anorexia	16.3
	Bulimia	20.5
	Bulimarexia	20.4
	Controls	21.1

Table 3
Weighted estimations of effect sizes for each clinical group and corresponding homogeneity tests

Clinical group	k	*g*	BI 95%	Q_W	P
Anorexia	**91**	**0.545**	**0.49/0.60**	**1826.94***	**0.001**
Bulimia	135	1.019	0.97/1.07	2270.32**	0.001
Bulimarexia	12	1.185	1.07/1.30	313.1***	0.001

* Excluding the study by Garner 1983b, in which two effect sizes were obtained.
** Excluding the studies by Gleghorn 1987 (BPSQ) and by Powers 1987 (BPSQ). One effect size was obtained in each study.
*** Excluding the study by Garner 1983b, in which two effect sizes were obtained.

the same way, we can attribute this different conclusion to the fact that in our study the statistical analysis is more powerful, on being applied to a larger number of studies.

Whole-body or body parts perceptual measures of distortion

With regard to the perceptual measures, they can be divided into two more categories that we should compare: techniques that assess the whole body and techniques that assess the body by parts. Table 4 shows the statistics for the anorexia and bulimia groups; the bulimarexia group is excluded due to the fact that, with only one effect size for each technique, it offers little information.

On evaluating the different types of perceptual measures used, we find that the body parts measures show a slightly stronger, but significant, effect than the whole-body measures [$d = .66$ and $d = .44$; $Q_B(1) = 14.44$; $p < .001$]. This result is not consistent with that of Cash & Deagle. Smeets and cols. (1997, 1998) also obtain a greater effect with whole-body measures, but it should be borne in mind that their study only included anorexia groups, and that Slade & Russell's (1973) technique, which measures the body by parts, shows exceptional results in their meta-analysis. The majority of our studies used mainly this technique in their evaluations of body image.

Body parts techniques obtain effect sizes statistically greater than whole-body techniques in the two clinical groups [anorexia, $Q_B(1) = 6.942$; $p < .008$; bulimia, $Q_B(1) = 8.61$; $p < .003$].

Global versus shape/weight attitudinal measures

Table 4 shows the mean of effect sizes of the attitudinal measures for the three diagnostic groups. In this analysis we made an explicit separation according to whether or not there were included two studies (Garner, 1983; Powers, 1987) whose values are quite different from the others, so that they can be considered as *outliers*. Whilst Cash & Deagle failed to obtain a significant effect for this categorical model, we obtained a significantly greater difference in global attitudinal measures than in shape/weight measures [$d = 1.06$ and $d = .26$; $Q_B(1) = 206.3$, $p < .001$].

Table 4

Analysis of the categorical models generated, for the clinical groups, as a function of type of measure, of whether the perceptual measure was carried out for the whole body or by parts, and of whether the attitudinal measure was of shape/weight or global

GROUP		**k**	***D***	**BI 95%**	**Qw**	**P**
			TYPE OF MEASURE			
Anorexia	Perceptual	48	0.384	0.301/0.466	251.1	0.001
	Attitudinal	41	0.673	0.597/ 0.748	1530.7	0.001
Bulimia	Perceptual	45	0.707	0.628/0.786	185.3	0.001
	Attitudinal	89	1.122	1.065/1.179	2135.5	0.001
Bulimarexia	Perceptual	2	1.344	1.026/1.662	0.91	0.634
	Attitudinal	8	1.162	1.04/1.28	311.1	0.001
			WHOLE BODY OR BY PARTS			
Anorexia	Parts	27	0.490	0.376/0.605	139.1	0.001
	Whole	21	0.269	0.150/0.387	105.1	0.001
Bulimia	Parts	25	0.835	0.721/0.950	142.3	0.001
	Whole	20	0.590	0.480/0.699	33.7	0.028
			ATTITUDE SHAPE/WEIGHT OR GLOBAL			
Anorexia	Shape/weight	12	0.549	0.376/0.722	388.1	.001
	Global	31	0.920	0.837/1.000	3128.5	.001
	Global*	_29_	_0.700*_	_0.618/0.786_	_1140.1_	_.001_
Bulimia	Shape/weight	12	0.079	-0.064/0.221	403.1	.001
	Global	78	1.259	1.198/1.321	1676.3	.001
	Global**	_77_	_1.180**_	_1.116/1.243_	_1591.0_	_.001_
Bulimarexia	Global	10	1.534	1.413/1.654	2492.6	.001
	Global*	_8_	_1.162*_		_311.1_	_.001_

* Excluding the study by Garner (1983b).
** Excluding the study by Powers (1987).

The bulimia group has a greater effect size in the global mode of assessment of cognitive aspects, this difference being statistically significant [$Q_B(1) = 224.5$; $p < .001$]. Furthermore, this group takes the lowest value for the measurement of the body according to weight and shape. The values for the anorexia group also show significant differences with respect to the different attitudinal measures [$Q_B(1) = 14.36$; $p < 0.001$].

From a more general perspective, of the four forms of evaluating body image, the global attitudinal measure is the category that shows the greatest effects, whilst the lowest value is produced on making an attitudinal evaluation of the weight and shape of the body.

Analysis of moderating variables

In order to evaluate the possible moderation of the eleven selected characteristics we applied variance and weighted regression analyses. As far as subject characteristics are concerned, neither age nor status of the controls modified the differences between the clinical samples and the controls, but there can be appreciated a significant association with BMI of the clinical group in the case of bulimia. Specifically, on carrying out a weighted regression analysis with BMI as predictor, on the effects obtained with bulimia groups, a significant effect was obtained [$Q_R(1) = 14.528$, $p < .001$; $Q_E(49) = 208.313$, $p < .001$; $R^2 = .065$]. The negative sign of the slope (-.14) indicates that in the bulimia patients, the higher the BMI, the smaller the difference in body-image distortion with respect to the controls. This appears to be a reasonable result, since the greater the degree of obesity, the more realistic (less distorted) the negative image perceived. Patients' status was relevant in its influence on the effect sizes, since inpatients presented higher values in body image (*d* = .80, 1.2 and 1.6 for the anorexia, bulimia and bulimarexia groups, respectively, all being statistically significant).

As regards the contextual variables, there is no significant association with the effect sizes. Publication date of the study differs according to clinical group, the 1980s for the anorexia group (*d* = 0.67) and the 1990s for bulimia and bulimarexia (*d* = 1.1 and 1.2). Comparable results are found for origin of the sample, where the European and Australian studies (*d* = .65 and .60) obtain greater and similar effect sizes for the anorexia group. In contrast, the bulimia and bulimarexia groups (*d* = 1.5 and 1.5) coincide in having a higher index in the European studies, but greater heterogeneity among the studies is implied.

With regard to the remaining methodological variables, the body part with greatest distortion differed between groups, with the waist and thighs obtaining the greatest effect size (*d* = 1.85) in bulimia and the thighs and face in anorexia (*d* = 1.4). These results contrast with our expectations, which situated the head as the most distorted part in the evaluation of body image. As regards type of diagnosis used, greater effect sizes were obtained with the DSM-III-R and the simultaneous use of various diagnostic criteria (*d* = 0.7, 1.3 and 1.2, respectively, for anorexia, bulimia and bulimarexia). These differences are significant for anorexia [$Q_B(5) = 104.3$; $p < .001$] and bulimia [$Q_B(3) = 284.67$; $p < .001$]. The affective instructions in the anorexia group produce the greatest differences between clinical and control subjects (*d* = 0.8), while for bulimia and bulimarexia it is ambiguous instructions that produce the greatest differences (*d* = .95 and 1.3); differences are statistically significant in all the clinical groups.

Special attention should be given to the model that distinguishes between studies in which the control group's behaviour is normal and those in which the controls present restrained eating behaviour. If the control group is restrained, the difference between it and the clinical group decreases, compares to the cases where it is not restrained [*d* = .55 and *d* = 1.6; $Q_B(1) = 72.8$, $p < .0001$]. If we take into account type of clinical group, we find that whilst there is scarcely any difference between the anorexia groups and the groups with restrained behaviour, this difference is large and significant in the case of the bulimia groups [*d* = .06 versus *d* = .98; $Q_B(1) = 34.3$, $p < .0001$].

Sources of bias and limitations of the meta-analysis

Our study did not include unpublished research, so that there is a potential lack of representativeness of our sample of studies (Rosenthal, 1979). On calculating the number of unpublished studies that would be necessary to alter the direction of our results (Rosenthal, 1991), we obtained a figure of 142,895 (1-a = 0.95). Consequently, we believe our study to be safe from this threat to its validity.

With regard to independence within a single study (Rosenthal & Rubin, 1986), we used up to twelve effect sizes of the same study. The decision to proceed in this way was made taking into account that the evaluation techniques varied, and that several studies had two control groups, one normal and the other with restrained behaviour. We do not believe, then, that this threat was realistically applicable to our meta-analysis.

CONCLUSIONS

The main conclusions of our meta-analysis are as follows:

(1) The fact of whether the *perceptual evaluation* is made through estimations of the whole body or by parts is important: greater effect sizes are obtained in the latter case. Therefore, results based on these two ways of evaluating image distortion should not be mixed.

(2) *Attitudinal measures* also present two modes of evaluation. The differences between patients and controls in relation to attitude towards global body image are greater than those in relation to attitude measured according to weight and body shape. These two modes of evaluation of cognitive-affective aspects of body image do indeed discriminate clinical groups (indices for the anorexia group referring to weight/shape and global, d = 0.54 and d = 0.92, whilst for the bulimia group the figures are d = 0.08 and d = 1.2, respectively). It can be concluded that body dissatisfaction in the clinical groups is greater when evaluation of their bodily appearance and attractiveness is involved.

The results of our meta-analysis differ markedly in some respects from those of the meta-analyses by Cash & Deagle (1997) and Smeets and cols. (1997, 1998). First of all, in the perceptual measures we obtained a difference between the anorexia and bulimia groups in the same direction as they did, but in our case it is statistically significant, whilst in their case it was not. Secondly, we obtained a difference between the whole body measures and those by body parts that was not obtained by them. We attributed this difference to the fact that in our meta-analysis there is less relative presence of anorexia groups and a larger number of studies that employ Slade & Russell's (1973) technique. Thirdly, whilst the other meta-analyses did not obtain a significant difference between the global attitudinal measures and those of shape/weight, we obtained a significantly greater effect for the global measures. If we distinguish between diagnostic groups, those of anorexia show a clearly greater effect in the global attitude measures, whilst those of bulimia do not show such a large effect. In our view, this means that in the anorexia groups the cognitive-affective factors have greater relative weight, since their physical appearance is characterized by slimness. In fact, in the meta-analysis by Smeets et al. (1998), in the anorexia group, lower BMI is associated with greater body overestimation, indicating a greater cognitive bias in the appreciation of their body. For their part, the bulimia groups have more justified body dissatisfaction if we bear in mind the model of slimness they wish to attain, and moreover, they have generally identified some parts of the body as those they would most like to reduce. Finally, the moderating variable relating to the control group was found to be highly relevant in the estimations of the effects. In fact, the control groups with restrained eating behaviour showed no differences in body dissatisfaction with respect to the anorexia groups, whilst in the case of normal controls a significant difference was appreciated. On considering bulimia groups we once again obtain a difference according to control group, though in this case there is indeed a significant difference even with respect to controls with restrained eating behaviour.

Our results show that attitudinal measures of body image have given rise to greater effect sizes than perceptual measures; therefore, it can be assumed that it is attitudes and beliefs about oneself that show the closest association with dissatisfaction.

Moreover, we should treat with caution the perceptual aspects of the so-called "distortion of the body image", since they are influenced by attitudinal aspects in relation to one's own body. Thus, the diagnostic criterion for anorexia expressed as "distortion of body figure perception" may be inappropriate for evaluating the concept of body image in eating disorders (Hsu, 1982; Reis et al., 1982). Specifically, this criterion, currently expressed in these terms, is not useful for discriminating clinical subjects from subjects from the normal population, since a large majority of adolescents and young adults perceive their weight or figure as unsatisfactory (Perpiñá & Baños, 1990; Paxton, 1993; Doll & Fairburn, 1998). In any case, we should like to insist on the fact that the diagnostic criterion expressed in perceptual terms as the evaluation of a young person within this disorder is not discriminative, and that the description of this distortion is merely a projection of dissatisfaction with the body, and not of a perceptual deficit. Therefore, it would be recommendable to modify this criterion in terms more related to attitude and beliefs about one's body.

Many homogeneity tests are significant, even when restricted to a particular clinical group and to a specific category of a categorical model. This indicates to us that there is still great variability in the results, due to one or more variables that we did not delimit clearly in the study, despite having analyzed a large quantity of variables. A tentative explanation may be found in the diagnostic criteria used in the selection of patients, not because of the criteria themselves but because of the patients, as many of them may be in a transitional stage of their illness.

Finally, one of our objectives was to suggest therapeutic lines based on the results obtained; the main one of these is the recommendation that future interventions and research should address cognitive-affective aspects, that is, patients' attitudes and beliefs about their body-image distortion. All eating disorder treatment should include specific attention to body image. This attention should focus on the following attitudinal dimensions: a) the evaluation of one's own body; b) a set of self-schemata (excess weight, body ideals); and c) affect towards one's own body in relation to specific situations. The duration of this intervention should be lengthy, since negative body image is probably one of the factors that initiates the disorder and prolongs it.

REFERENCES

(Research reports included in the meta-analysis are marked with an asterisk)

Alleback, P., Hallberg, D. & Espmark, S. (1976). Body image-An apparatus for measuring disturbances in estimation of size and shape. *Journal of Psychosomatic Research*, 20, 583-589.

American Psychiatric Association (1979). *Diagnostic and statistical manual of mental disorders (3rd edition).* Washington: Author.

American Psychiatric Association (1987). *Diagnostic and statistical manual of mental disorders (3rd edition rev.).* Washington: Author.

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders (4th edition).* Washington: Author.

Askevold, R. (1975). Measuring body image: Preliminary report on a new method. *Psychotherapy and Psychosomatics*, 26, 71-77.

*Ben-Tovim, D. I., Whitehead, J. & Crisp, A. H. (1979). A controlled study of the perception of body width in anorexia nervosa. *Journal of Psychosomatic Research*, 23, 267-272.

*Berman, K., Lam, R.W. & Goldner, E.M. (1993). Eating attitudes in seasonal affective disorder and bulimia nervosa. *Journal of Affective Didorders*, 29, 219-225.

*Birtchnell, S. A., Lacey, J. H. & Harte, A. (1985). Body image distortion in bulimia nervosa. *British J. of Psychiatry*, 147, 408-412.

*Bowden, P. K., Touyz, S.W., Rodriguez, P.J., Hensley, R. & Beaumont, P.J.V. (1989). Distorting patient or distorting instrument? *British J. of Psychiatry*, 155, 196-201.

*Bunnell, D. W., Cooper, P. J., Hertz, S. & Shenker, I. R. (1992). Body shape concerns among adolescents. *International Journal of Eating Disorders*, 11, 79-83.

*Button, E. J., Francella, F. & Slade, P. D. (1977). A reappraisal of body perception disturbance in anorexia nervosa. *Psychological Medicine*, 7, 235-243.

Bruch, H.(1962). Perceptual and conceptual disturbance in anorexia nervosa. *Psychosomatic Medicine*, 24, 187-194.

Cash, T. F. (1995). *What do you see when you look in the mirror?* New York, Bantam Books.

Cash, T. F. & Deagle, E. A. (1997). The nature and extent of body image disturbances in anorexia nervosa and bulimia nervosa: a meta-analysis. *International Journal of Eating Disorders*, 22, 107-125.

*Casper, R. C., Halmi, K. A., Goldberg, S. C., Eckert, E. D. & Davis, J.M. (1979). Disturbances in body image estimation as related to other characteristics and outcome in anorexia nervosa. *British Journal of Psychiatry*, 134, 60-66.

*Collins, J. K. (1986). The objetive measurement of body image using a video technique: Reliability and validity studies. *British Journal of Psychology*, 77, 199-205.

*Collins, J.K. (1987). Methodology for the objective measurement of body image. *International Journal of Eating Disorders*, 6, 393-399.

*Collins,J.K., Beumont, P. J., Touyz, S. W., Krass, J., Thompson, P. & Philips, T. (1987^{a}). Accuracy of body image with varying degrees of information about the face and body contours. *International Journal of Eating Disorders*, 6, 67-73.

*Collins, J.K., Beumont, P. J., Touyz, S. W., Krass, J., Thompson, P. & Philips, T. (1987b). Variability in body shape perception in anorexic, bulimic, obese, and control subjects. *International Journal of Eating Disorders*, 6, 636-638.

*Cooper, P. J., Taylor, M. J., Cooper, Z., & Fairburn, C. G. (1987). The development and validation of the Body Shape Questionaire. *International Journal of Eating Disorders*, 6, 485-494.

*Cooper, Z., Cooper, P. J. & Fairburn, C. G. (1989). The validity of the Eating Disorder Examination and its subscales. *British Journal of Psychiatry*, 154, 807-812.

*Cooper, P.J. & Taylor, M.J. (1988). Body image disturbance in bulimia nervosa. *British J. of Psychiatry*, 153 (suppl. 2), 32-36.

*Cooper, M.J. & Fairburn, C.G. (1992). Thoughts about eating, weight and shape in anorexia nervosa and bulimia nervosa. *Behaviour Res. Therapy,* 5, 501-511.

*Cooper, M.J., Anastasiades, P. & Fairburn, C.G. (1992). Selective processing of eating-, shape-, and weight-related words in persons with bulimia nervosa. *Journal of Abnormal Psychology*, 2, 352-355.

*Dacey, C.M., Nelson, W. M., Clark, V.F. & Aikman, K.G. (1981). Bulimia and Body Image dissatisfaction. *Child Psychiatry and Human Development*, 21, 179-184.

*Doll, H.A., & Fairburn, C.G. (1998). Heightened accuracy of self-reported weight in bulimia nervosa: A useful cognitive “distortion”. *International Journal of Eating Disorders* 24, 267-273.

Feighner, J. P., Robins, E., Guze, S, Woodruff, R. A. Winokur, G. & Muñoz, R. (1972). Diagnostic criteria for use in psychiatric research. *Archives General Psychiatry*, 26, 57-63.

*Fichter, M.M., Meister, I. & Koch, H. I. (1986). The measurement of body image disturbance in anorexia nervosa: experimental comparison of different methods. *British Journal of Psychiatry*, 148, 453-461.

*Franzen, U., Florin, I., Schneider, S. & Meier, M. (1988). Distorted body image in bulimic women. *Journal of Psychosomatic Research*, 32, 445-450.

*Franzoi, S. L. & Shields, S. A. (1984). The Body Esteem Scale: Multidimensional structure and sex differences in a college population. *Journal of Personality Assessment*, 48, 173-178.

*Freeman, R., Touyz, S., Sara, G., Rennie, C., Gordon, E. & Beumont, P. (1991). In the eye of the beholder: processing body shape information in anorexic and bulimic patients. International Journal of Eating Disorders, 6, 709-714.

*Freeman, R. J., Thomas, C. D., Solyom, L. & Miles, J. E. (1983). Body image disturbance in anorexia nervosa: A reexamination and a new technique. In P. L. Darby, P. E. Garfinkel & D. M. Garner (Eds.), *Anorexia nervosa: New developments in research* (pag. 117-127). New York: Alan R. Liss.

*Freeman, R. J., Thomas, C. D., Solyom, L. & Hunter, M. A. (1984). A modified video camera for measuring body image distortion: Technical description and reliability. *Psychological Medicine*, 14, 411-416.

*Freeman, R. J., Thomas, C. D., Solyom, L. & Koopman, R. F. (1985). Clinical and Personality correlates of body size overestimation in anorexia nervosa and bulimia nervosa. *International Journal of Eating Disorders*, 4, 439-456.

*Friess, H. (1977). Studies on secondary amenorrhea, anoretic behavior, and body image perception. In R. A. Vigersky. (Eds.), *Anorexia nervosa* (pag. 173-176). New York: Raven Press.

*Gardner, R. M. & Moncrieff, C. (1988). Body image distortion in anorexics as a non-sensory phenomenon: a signal detection approach. *Journal of Clinical Psychology*, 44, 101-107.

*Garfinkel, P. E., Garner, D. M., Rose, J., Darby, P. L., Brandes, J. S., O’Hanlon, J. & Walsh, N. (1983). A comparison of characteristics in the families of patients with anorexia nervosa and controls. *Psychological Medicine*, 13, 821-828.

*Garfinkel, P. E., Moldofsky, H., Garner, D. M., Stancer, H. C. & Coscina, D. V. (1978). Body awareness in anorexia nervosa: Disturbances in “body image” and “satiety”. *Psychosomatic Medicine*, 40, 487-498.

*Garner, D.M., Olmsted, M.P. & Garfinkel, P.E. (1983). Does anorexia nervosa occur on a continuum? *International Journal of Eating Disorders*, 3, 11-20.

*Garner, D. M., Garfinkel, P. E., Stancer, H. C. & Moldofsky, H. (1976). Body image disturbance in anorexia nervosa and obesity. *Psychosomatic Medicine*, 38, 327-336.

*Garner, D. M., Garner, M. V. & Van Egeren, L. F. (1992). Body dissatisfaction adjusted for weight: The Body Illusion Index. *International Journal of Eating Disorders*, 11, 263-271.

*Garner, D. M., Olmsted, M. P. & Polivy, J. (1983^a). Comparison between weight preoccupied women and anorexia nervosa. *Psychosomatic Medicine*, 46, 255-266.

*Garner, D. M., Olmsted, M. P. & Polivy, J. (1983b).

Development and validation of a multidimensional eating disorder inventory for anorexia nervosa and bulimia. *International Journal of Eating Disorders,* 2, 15-34.

Glass, G.V. (1976). Primary, secondary and meta-analysis of research. *Educational Researcher,* 10, 3-8.

*Gleghorn, A. A., Penner, L. A., Powers, P. S. & Schulman, R. (1987). The psychometric properties of several measures of body image. *Journal of Psychopathology and Behavioural Assessment,* 9, 203-218.

Gómez, J. (1987). *Meta-análisis*. Promociones y Publicaciones Universitarias, S.A., Barcelona.

*Gross, J, Rosen, J. C., Leitenberg, H. & Willmuth, M. E. (1986). Vadility of the Eating Attitude Test and the Eating Disorder Inventory for bulimia nervosa. *Journal of Consulting and Clinical Psychology*, 54, 875-876.

*Hadigan, C. M. & Walsh, B. T. (1991). Body shape concerns in bulimia nervosa. *International Journal of Eating Disorders,* 10, 323-331.

Halmi, K. A., Goldberg, S. C. & Cunningham, S. (1977). Perceptual distortion of body image in adolescent girls. *Psychological Medicine*, 7, 253-257.

*Hamilton, K. & Waller, G. (1993). Media influences on body size estimation in anorexia and bulimia. *British Journal of Psychiatry*, 162, 837-840.

Hedges, L.V. (1981). Distribution theory for Glass' estimator of effect size and related estimators. *Journal of Educational Statistics,* 6, 197-228.

Hedges, L.V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando: Academic Press.

*Horne, R. L., Van Vactor, J. C. & Emerson, S. (1991). Disturbed body image in patients with eating disorders. *American Journal of Psychiatry,* 148, 211-215.

Hsu, L. K. (1982). Is there a disturbance in body image in anorexia nervosa? *Journal of Nervous and Mental Disease*, 170, 305-307.

*Huon, G. F. & Brown, L. B. (1986). Body image in anorexia nervosa and bulimia nervosa. *International Journal of Eating Disorders,* 5, 421-439.

*Huon, G. F. & Brown, L. B. (1989). Assessing bulimics' dissatisfaction with their body. *British Journal of Clinical Psychology,* 28, 283-284.

Johnson, B. T. (1989). *DSTAT: software for the meta-analytic review of research literature*. Hillsdale, N. J. Erlbaum. (Revised 1993).

*Joiner, T.E., Schmidt, N.B. & Wonderlich, S.A. (1997). Global self-esteem as contingent on body satisfaction among patients with bulimia nervosa: Lack of diagnostic specificity? *International Journal of Eating Disorders* 1, 67-76.

*Joiner, T.E., Wonderlich, S.A., Metalsky, G.I. & Schmidt, N.B. (1995). Body dissatisfaction: A feature of bulimia, depression, or both? *Journal of Social and Clinical Psychology*, 4, 339-355.

*Kalliopuska, M. (1982). Body-image disturbances in patients with anorexia nervosa. *Psychological Reports*, 51, 715-722.

*Katzman, M.A. & Wolchik, S.A. (1984). Bulimia and binge eating in college women: A comparison of personality and behavioral characteristics. *Journal of Consulting and Clinical Psychology,* 3, 423-428.

Kolb, L.K. (1975). Disturbance of the body image. *Psychosomatic Medicine*, 33, 810-837.

*Laessle, R. G., Tuschl, R. J., Waadt, S. & Pirke, K. M. (1989). The specific psychopathology of bulimia nervosa: A comparison with restrained and unrestrained (normal) eaters. *Journal of Consulting and Clinical Psychology,* 57, 772-775.

*Leon, G. R., Lucas, A. R., Colligan, R. C., Ferdinande, R. J. & Kamp, J. (1985). Body image, sexual attitudes and family interaction patterns in anorexia nervosa. *Journal of Abnormal Child Psychology,* 13, 245-258.

*Lindholm, L. & Wilson, G. T. (1988). Body image assessment in patients with bulimia nervosa and normal controls. *International Journal of Eating Disorders*, 7, 527-639.

*Long, C.G., Hinton, C. & Gillespie, N.K. (1994). Selective processing of food and body size words: Application of the Stroop test with obese restrained eaters, anorexics, and normals. *International Journal of Eating Disorders* 3, 279-283.

*Lovell, D.M. & Williams, M. G. (1997). Selective processing of shape-related words in women with eating disorders, and those who have recovered. *British Journal of Clinical Psychology*, 36, 421-432.

*McKenzie, S. J., Williamson, D. A. & Cubic, B. A. (1993). Stable and reactive body image disturbance in bulimia nervosa. *Behavior Therapy,* 24, 195-207.

*Meerman, R. (1983). Experimental investigation of disturbance in body image estimation in anorexia nervosa patients, and ballet and gymnastic pupils. *International Journal of Eating Disorders*, 2, 91-100.

*Mizes, J. S. (1988). Personality characteristics of bulimic and non-eating disordered female controls: A cognitive behavioral perspective. *International Journal of Eating Disorders*, 7, 541-550.

*Mizes, J. S. (1992). The Body Image Detection Device versus subjective measures of weight dissatisfaction: A validity comparison. *Addictive Behaviors*, 17, 125-136.

*Nelson, C. L. & Gidycz, C. A. (1993). A comparison of body image perception in bulimics, restrainers and normal women: An extension of previous findings. *Addictive Behaviors*, 18, 503-509.

*Norris, D. L. (1984). The effects of mirror confrontation on self-estimation of body dimensions in anorexia nervosa, bulimia and two control groups. *Psychological Medicine*, 14, 835-842.

*Nudelman, S., Rosen, J. C. & Leitenberg, H. (1988). The similarities in eating attitudes, body image distortion, depression and self-esteem between high-intensity male runners and women with bulimia nervosa. *International Journal of Eating Disorders*, 7, 625-634.

Paxton, S.J. (1993). A prevention program for disturbed eating and body dissatisfaction in adolescent girls: a 1 year follow-up. *Health Education Research*, 8, 43-51.

Perpiñá, C. & Baños, R.M. (1990). Distorsión de la imagen corporal: Un estudio en adolescentes. *Anales de Psicología*, 1, 1-9.

*Pierloot, R. A. & Houben, M. F. (1978). Estimation of body dimensions in anorexia nervosa. *Psychological Medicine*, 8, 317-324.

*Powers, P. S., Schulman, R. G., Gleghorn, A. A. & Prange, M. E. (1987). Perceptual and cognitive abnormalities in bulimia. *American Journal of Psychiatry*, 144, 1456-1460.

*Probst, M., Van Coppenolle, H., Vandereycken, W. & Goris, M. (1992). Body image assessment in anorexia nervosa patients and university students by means of video distortion: a realiability study. *Journal of Psychosomatic Research*, 36, 89-97.

*Probst, M., Vandereycken, W., Van Coppenolle, H. & Vanderlinden, J. (1995). The body attitude test for patients with an eating disorder: psychometric characteristics of a new questionnaire. *Eating Disorders*, 3, 133-144.

*Probst, M., Vandereycken, W. Van Coppenolle, H. & Pieters, G. (1995). Body size estimation in eating disorder patients: testing the video distortion method on a life-size screen. *Behaviour Res. Therapy*, 8, 985-990.

*Proctor, L. & Morley, S. (1986). “Demand characteristic” in body size estimation in anorexia nervosa. *British Journal of Psychiatry*, 149, 113-118.

Reis, H.T. Wheeler, L. Spiegel, N., Kernis, M., Nezlek, J. & Perri, M. (1982). Physical attractiveness in social interaction, II: Why does appearance affect social experience? *Journal of Personality and Social Psychology*, 43, 979-996.

*Rodríguez, M.C. (1997). Distorsión de la imagen corporal como factor de riesgo para la anorexia nerviosa: un estudio comparativo. *Anales de Psiquiatría*, 8, 327-331.

*Rosen, J.C. & Ramirez, E. (1998). A comparison of eating disorders and body dysmorphic disorder on body image and psychological adjustment. *Journal of Psychosomatic Research*, ?, 441-449.

*Rosen J.C., Reiter, J. & Orosan, P. (1995). Assessment of body image in eating disorders with the body dysmorphic disorder examination. *Behavior Research Therapy,* 1, 77-84.

*Rosen, J. C., Srebnik, D., Saltzberg, M. & Wendt, S. (1991). Development of a Body Image Avoidance Questionnaire. *Psychological Assessment*, 3, 32-37.

Rosenthal, R. (1979). The “file drawer problem” and tolerance for null results. *Psychological Bulletin*, 86, 638-641.

Rosenthal, R. & Rubin, R.D. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin*, 3, 400-406.

Rosenthal, R. (1991). *Meta-analytic procedures for social research (ed. Rev.).* Newbury Park, C.A.: Sage.

*Ruff, G.A. & Barrios, B.A. (1986). Realistic assessment of body image. *Behavioral Assessment*, 8, 237-252.

Russell, G. (1979). Bulimia nervosa: an ominous variant of anorexia nervosa. *Psychological Medicine*, 9, 429-448.

Sánchez, J. & Ato, M. (1989). Meta-análisis: Una alternativa metodológica a las revisiones de la investigación. In Arnau, J. & Carpintero, H. (Coords.) *Tratado de Psicología General I: Historia, teoría y método* (pp. 617-669). Madrid: Alhambra.

Sánchez-Carracedo, D. & Saldaña, C. (1998). Evaluación de los hábitos alimentarios en adolescentes con diferentes índices de masa corporal. *Psicothema*, 10(2), 281-292.

*Schulman, R. G., Kinder, B. N., Powers, P. S. & Gleghorn, A. (1986). The development of a scale to measure cognitive distortion in bulimia. *Journal of Personality Assessment*, 50, 630-639.

*Slade, P. D., Dewey, M. E., Newton, T. & Brodie, D. A. (1990). Development and preliminary validation of the Body Satisfaction Scale (BSS). *Psychology and Health*, 4, 213-220.

*Slade, P. D. & Russell, G. F. M. (1973^a). Awareness of body dimensions in anorexia nervosa: Cross-sectional and longitudinal studies. *Psychological Medicine*, 3, 188-199.

Smith, M.L. & Glass, G.V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychology*, 32, 752-760.

Steinhausen, H. C. & Glanville, K. (1983). A long term follow up of adolescents' anorexia nervosa. *Acta Psychiatrica Scandinavica*, 68, 1-10.

*Striegel-Moore, R. H., Silberstein, L. R. & Rodin, J. (1993). The social self in bulimia nervosa:. Public self-consciousness, social anxiety and perceived fraudulence. *Journal of Abnormal Psychology*, 102, 297-303.

*Strober, M., Goldenburg, I., Green, J. & Saxon, J. (1979). Body image disturbance in anorexia nervosa during acute and recuperative phase. *Psychological Medicine*, 9, 201-204.

*Sunday, S. R., Halmi, K. A., Werdann, L. & Levey, C. (1992). Comparison of body size estimation and eating disorder inventory scores in anorexia and bulimia patients with obese, and restrained and unrestrained controls. *International Journal of Eating Disorders*, 11, 133-149.

*Tadai, T., Kanai, H., Nakamura,M., Nakajima, T., Fujita, M. & Nakai, Y. (1994). Body image changes in adolescents II comparison among patients with eating disorders and controls with thin, normal and obese body shapes. *The Japanese J. of Psychiatry and Neurology*, 48, 541-545.

*Thompson, J.K., Berland, N.W., Linton, P.H. & Weinsier, R.L.(1986). Utilization of a self-adjusting light beam in the objective assessment of body distortion in seven eating disorder groups. *International Journal of Eating Disorders* 1, 113-120.

*Touyz, S. W., Beaumont, P. J. V., Collins, J. K., McCabe, M. & Jupp, J. (1984). Body shape perception and its disturbance in anorexia nervosa. *British Journal of Psychiatry*, 144, 167-171.

*Whitehouse, A. M., Freeman, C. P. L. & Annandale, A. (1986). Body size estimation in bulimia. *British Journal of Psychiatry*, 149, 98-103.

*Whitehouse, A. M., Freeman, C. P. L. & Annandale, A. (1988). Body size estimation in anorexia nervosa. *British Jouranl of Psychiatry*, 153 (supple. 2), 23-26.

*Williamson, D. A., Cubic, B. A. & Gleaves, D. H. (1993). Equivalence of body image disturbance in anorexia and bulimia nervosa. *Journal of Abnormal Psychology*, 102, 177-180.

*Williamson, D. A., Davis, C. J., Bennett, S. M., Goreczny, A. J. & Gleaves, D. H. (1989). Development of a simple procedure for assessing body image disturbances. *Behavioral Assessment*, 11, 433-446.

*Williamson, D. A., Davis, C. J., Goreczny, A. J. & Blouin, D. C. (1989). Body image disturbances in bulimia nervosa: influences of actual body size. *Journal of Abnormal Psychology*, 98, 97-99.

*Williamson, D. A., Gleaves, D. H., Watkins, P. C. & Schlundt, D. G. (1993). Validation of self ideal body size discrepancy as a measure of body dissatisfaction. *Journal of Psychopathology and Behavioral Assessment*, 15, 57-68.

*Williamson, D. A., Kelly, M. L., Davis, C. F., Ruggerio, L. & Blouin, D. C. (1985). Psychopathology of eating disorders: A controlled comparison of bulimic, obese and normal subjects. *Journal of Consulting and Clinical Psychology*, 53, 162-166.

*Willmuth, M. E., Leitenberg, H., Rosen, J. C. & Cado, S. (1988). A comparison of purging and non purging normal weight bulimics. *International Journal of Eating Disorders*, 7, 825-835.

*Willmuth, M. E., Leitenberg, H., Rosen, J. C., Fondacaro, K. M. & Gross, J. (1985). Body size distortion and bulimia nervosa. *International Journal of Eating Disorders*, 4, 71-78.

*Willson, G.T. & Smith, D. (1989). Assessment of bulimia nervosa: An evaluation of the Eating Disorders Examination. *International Journal of Eating Disorders*, 8, 173-179.

*Wooley, O. W. & Rull, S. (1991). The Colour-A-Person Body Dissatisfaction Test: Stability, internal consistency, validity and factor structure. *Journal of Personality Assessment*, 56, 395-413.

Psychology in Spain, 2002, Vol. 6. No 1, 96-101

PROGRAMME FOR THE IMPROVEMENT OF METAMEMORY IN PEOPLE WITH MEDIUM AND MILD MENTAL RETARDTION

Lourdes Martínez Pérez* and Emilio García García**
**Colegio Virgen de Lourdes, Majadahonda, Madrid. ** Complutense University of Madrid*

La metamemoria es el conocimiento de nuestra memoria y de todo aquello que es relevante para el registro, almacenamiento y recuperación de la información. Mediante ella somos capaces de detectar si una situación requiere un esfuerzo mayor o menor, qué estrategias son las más adecuadas en función de la tarea, nuestras características cognitivas y el contexto en el que estamos. La metamemoria se desarrolla con la edad, y nos permite una sofisticación de las estrategias para el aprendizaje. Las investigaciones en este campo indican que a mayor metamemoria mejores ejecuciones en memoria. Los déficits de memoria en deficiencia mental se atribuyen, entre otras cosas, a los pobres conocimientos que los deficientes mentales poseen sobre su memoria. Esta es la razón por la que nosotros hemos diseñado este programa. Los resultados confirman la efectividad del mismo, ya que gracias a él, los deficientes mentales son capaces de aumentar sus conocimientos sobre su memoria.

Metamemory is knowledge about our memory and about everything that is relevant for the recording, storing and recovery of information. It enables us to detect whether a situation requires greater or lesser effort, and what strategies are most suitable according to the task, our cognitive characteristics and the context. Metamemory develops with age, allowing us to increase the sophistication of our learning strategies. Research in this field indicates that the more developed the metamemory, the better the memory performance. Memory deficits in people with mental retardation are attributed to, among other factors, their poor knowledge about their memory. It is this that motivated us to design the programme. The results confirm the effectiveness of the programme, which indeed helps this population to increase knowledge about their memory.

People with mental retardation have serious difficulties with the majority of intentional memory tasks, the causes of these difficulties being cited as brain damage, metacognitive deficits or a poor knowledge base.

The aim of this work is to check whether people with mental retardation are capable of benefiting from training in metamemory and, if they are, to determine whether they are capable of maintaining over time the knowledge acquired. This research was carried out with a view to seeking possible forms of intervention to improve learning in people with mental retardation.

Metamemory is the knowledge we have about memory in general and about our own memory in particular. Flavell and Wellman (1977), define it as our knowledge of the memory and of all that is relevant for the recording, storing and recovery of information. This knowledge permits us to put into practice a series of strategies for solving our memory problems in any everyday situation (García, 1977).

The majority of studies in people with mental retardation suggest that many of their memory problems derive from their poor knowledge about their memory, which makes it difficult for them to use strategies correctly and to generalize the strategies learned (Borkowski and Wanschura, 1974; Borkowski, Millstead and Hale, 1988; Bellinger, Borkowski, Turner and Hale, 1995; Turner, Haley and Borkowski, 1996; Valkil, Shelef-Reshef and Levy-Shiff, 1997). Other research indicates deficits in control as the cause of their difficulties in memory tasks, since they are incapable of planning, monitoring or evaluating their performance (Sternberg, 1985; Borkowski, Reid and Kurtz, 1984; Flavell, 1987; Butterfield and Belmont, 1977; Brown, 1978).

Our objective is to design a programme for improving the knowledge of people with mental retardation in relation to memory, which covers aspects of both knowledge and control. The specific aims of the present work can be summarized as follows: a) To detect the know-

---

The original Spanish version of this paper has been previously published in *Psicología Educativa*, 2001, Vol.. 7. No 1, 63-74

....

Correspondence concerning this article should be addressed to Lourdes Martínez Pérez. C/ Luis Velez de Guevara, 9-2º-14. 28012 Madrid. E-mail: lmartinez@correo.cop.es ledge of people with mild and medium mental retardation about their memory; b) To apply a programme for increasing this knowledge; C) To check the effectiveness of this programme.

METHOD

Subjects

The sample used for the research was made up of 28 people with mild and medium mental retardation, with a mean I.Q. of 52, assessed using the WISC-R. Chronological age ranged from 13 to 17 years, with a mean of 15 years 3 months. All were from the same institution. For the assignment of the experimental and control groups we took as a reference the scores obtained in Belmont and Borkowski's (1988) Metamemory Battery.

Hypotheses

1- People with mental retardation have a poor metamemory.
2- They can benefit from training in metamemory.

The training programme

Our training programme combines general and specific strategies and knowledge, since it appears that such programmes are those that produce most benefits in people with mental retardation.

Our instructional approach took into account the principles of Brown and Palincsar (1982) for teaching any strategy and the metacognitive training patterns of Osman and Hannafin (1992) and Borkowski (1992). We employed a discriminative phase with regard to the situations in which it is and is not appropriate to apply a given strategy (Mayor, 1988; Mayor and Sainz, 1988) and provided feedback on the choice and execution of strategies (Campione, Brown and Ferrara, 1982). Finally, we tried to ensure that the adolescents assessed their performance, as this increases their potential for transfer (Cox, 1994).

The general objective of the programme is that the participants acquire clear knowledge of what memory is (general metacognitive knowledge) and what they can do to improve their memory performance, and they are taught specific strategies (repetition and organization) and general ones (planning and control). This general aim can be subdivided as follows: a)- To help the adolescents to identify and define problems, training planning as a general and highly generalizable strategy; b)- To teach them to consider effective strategies for the solution of a problem; c)- To help them to assimilate self-assessment techniques, assessing their personal work and their weak and strong points; with this objective we worked on the knowledge variable person and the control variable checking; and d)- To help them to locate connections between the tasks trained and everyday life, working on generalization.

The programme consisted of 40 one-hour sessions, which were applied from October until April. The experimental group was divided into two groups for the training, which was carried out during school time. The application of the programme was structured in four stages: Introduction, Teaching of metacognitive variables of knowledge and control, Conceptualization (planning and checking), and Cognitive strategies of repetition and categorization. These were followed by a Combined phase.

In the *Introduction stage* we told the adolescents about the characteristics of the work they would be doing and the goals proposed, and provided a general outline of the programme.

The *Metacognitive knowledge variables stage* was focused on defining among everyone what memory is and what its uses are, detecting the participants' strong and weak points, defining what a strategy is, identifying and being aware of the strategies they used, judging whether strategies helped us to recall things better, and detecting which type of material was easiest to recall and which was most difficult, and why. This stage included a modelling phase, a guided phase and an individual phase.

In the *Conceptualization stage* we worked on concepts and categories, which are fundamental for organization strategies.

In the *Control variables stage* we focused on identifying and defining memory problems themselves, seeking different ways of solving them according to the participants' personal characteristics and those of the task. Subsequently, we assessed the results obtained, both individually and in groups.

In the final *Combined phase* we integrated all the stages.

All the sessions except that of the presentation had the following structure: 1- Activation of previous knowledge, with a brief summary of the last session; 2- Performing of exercises corresponding to that session; and 3- Session ends with a synthesis of it.

Example of a person variables session

The objectives of these sessions is that the participants become aware of their strong and weak points in relation to memory and of the spontaneous strategies they use. The session begins with the activation of previous knowledge, stressing the concept of memory (we should point out that in the presentation sessions we had already defined among everyone what memory is and what its uses are). In the modelling phases the researcher mentions two activities that s/he recalls with no problem, and explains why s/he has no difficulty. S/he then says the following to the adolescents: " *Now you are going to tell us two things that you remember without any difficulty, and just as I did, you must explain to us why it's no problem"*.

The adolescents write their replies in a notebook. When all of them have finished, they must read them out to the class and demonstrate that they can easily recall that task or situation. Subsequently, they must explain why they have no difficulty to recall them. It is important to listen to all the replies given by the participants. It should be borne in mind that in principle there are no right or wrong responses, and all are noted down on the blackboard. It is most important that the researcher makes no judgement about any reply, since the aim of the session is for all the participants to "become aware" of their strong and weak points. We believe that if we do not take note of all the replies given by the adolescents, this may cause inhibition in those with the greatest difficulties. Subsequently, the same procedure is carried out for two tasks that the adolescents find it difficult to remember. It may be that in these initial sessions the replies given by the adolescents are not appropriate, and in this case we can suggest everyday situations in which they do not normally have problems and allow each to respond in turn.

Table 1
Means and standard deviations of the experimental and control groups at the three measurement points

VARIABLES	**Experimental group**			**Control group**		
	M1	**M2**	**M3**	**M1**	**M2**	**M3**
Total score in metamemory	10.00 (2.45)	19.46 (3.97)	20.08 (5.29)	9.71 (2.18)	12.42 (5.15)	11.29 (3.87)
Organized list	1.00 (1.52)	4.00 (1.63)	4.46 (2.18)	0.57 (1.22)	1.86 (1.66)	0.71 (1.68)
Preparation of the object	1.50 (0.86)	2.54 (1.45)	3.69 (1.50)	1.93 (0.99)	2.56 (1.49)	2.42 (1.30)
Associated pairs	0.79 (1.67)	4.77 (2.29)	3.69 (2.69)	0.43 (0.85)	1.57 (2.50)	1.28 (2.16)
Circular recall	0.14 (0.54)	0.62 (1.75)	0.61 (0.96)	0.00 (0.00)	0.14 (0.54)	0.14 (0.53)
Rating of memory	6.57 (1.02)	7.54 (1.20)	7.69 (1.84)	6.64 (1.21)	6.50 (1.10)	6.71 (2.12)

Example of a combined session in which we unite the task variables, the repetition strategies and the metacognitive variables of checking and control.

The aim of these sessions is that the participants identify which material is easiest to remember and why. We should stress that known material is easier than unknown material, that remembering a little material is simpler than remembering a lot of material and that organization of the material favours recall. We teach the use of the cumulative repetition strategy. Our aim is for the adolescents to detect the nature of the problem and to check their performance. It should be borne in mind that all of these variables must initially be worked on individually. The material that can be used consists of two lists with six words in each, one with closely related words, such as animals, and one with non-related words.

In the modelling phase the researcher explains the entire process in the following way: "If I have to choose one of these two word lists to learn, let's see which one is easier for me. In order to find this out, I have to read them" (s/he reads them aloud). On reading the first list s/he says aloud "it has six words and they're all animals". The second list is then read aloud, and s/he says: "it also has six words, but these can't be put into a group, because one is a piece of furniture, one is an item of clothing, and so on." The researcher then talks about the comparison of the two lists, deciding that the first is simpler, since the words are related. Next, she says: "Right, the following step is to learn them and I'm going to repeat them; I know they're animals and that I have to repeat 6 words." The repetition is made aloud. "Once we think we've learned them properly we check our learning and decide that we know them."

This is followed by the guided phase, in which the adolescents are helped with carrying out the exercise. A series of questions is provided to guide them: "What do

Table 2
Significant effects in the different variables analyzed

VARIABLES	**F. Group DF (1, 25)**	**F. Change DF (2,50)**	**F. Interaction DF (2,50)**	**Contrast 1**	**Contrast 2**
Total score in metamemory	17.30***	45.45***	19.09***	67.07***	37.14***
Organized list	18.78***	19.59***	10.08***	15.07***	0.01NS
Preparation of object	0.75NS	11.77***	4.81*	6.82*	0.19NS
Associated pairs	9.80**	17.21***	5.14**	6.9*	6.40*
Circular recall	2.75NS	1.26NS	0.35NS	1.32NS	0.19NS
Rating of memory	3.52NS	1.14NS	1.19NS	3.49NS	1.79NS

NS: non-significant *p< 0.05 **p< 0.01 ***p<0.001

we have to do?", "What are we going to do first?", "Which list is easier?", "Why?", "How do I have to learn it?", "Am I sure I know it now?", "Why?".

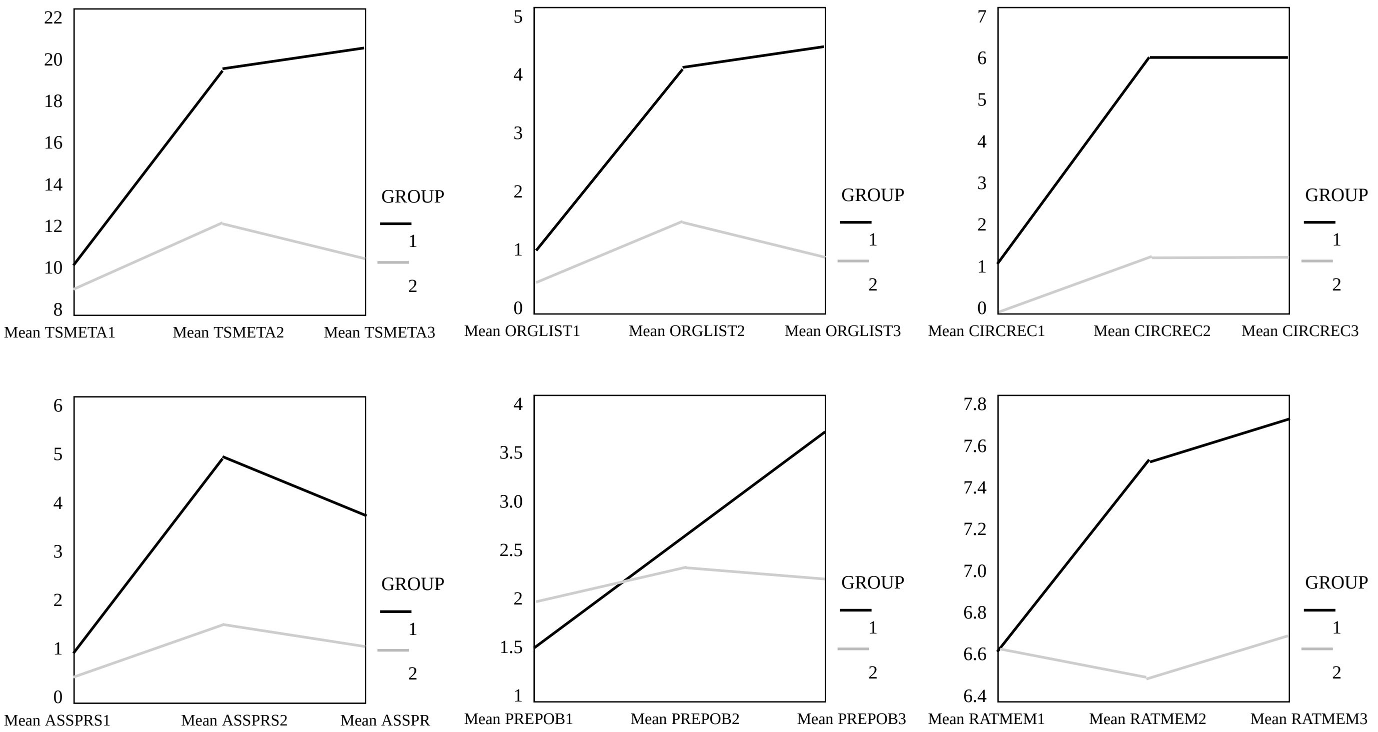
In the individual phase the participants carry out the task alone, and subsequently tell the group what they have done. The assessments are made by the group.

RESULTS

The statistical analysis used was that of the General Linear Model of Repeated Measures, Type III. Treatment acted as the between-subjects variable, dividing the population into experimental group and control groups. As within-subjects variable we used the measures taken in the Metamemory Battery before, after and in the maintenance phase. For the effects of the research we took into account only the analyses corresponding to the interaction between the between-subjects variable and the within-subjects variable, since our objective was to determine the effects of the training in the experimental group. Table 1 shows performance at the three measurement points. In order to determine the point at which the differences occur we carried out a "post-hoc" contrast, specifically a deviation contrast, which is a within-subjects contrast.

The results of the repeated-measures analysis (Table 2) indicate significant effects in the following variables: Total score in metamemory, F(2,50)=19.09, $p< 0.001$, and the subtests Organized list, F(2,50)=10.08, $p<0.001$, Preparation of object, F(2,50)=4.81, $p<0.05$ and Associated pairs, F(2,50)=5.14, $p<0.001$. The contrasts indicate the point at which the changes have occurred. We applied "post-hoc" contrasts, observing that in the variable "Total score in metamemory" the change occurred between the scores pre/maintenance, Contrast 1, t=67.07 $p< 0.001$ and post/maintenance, Contrast 2, t=37.14, $p<0.001$. In the "Associated pairs" variable the changes also took place in the pre/maintenance phase, Contrast 1, t=6.9, $p< 0.05$ and the post/maintenance phase, Contrast 2, t=6.40, $p< 0.05$. In the variables "Organized list" and "Preparation of the object" these changes were only observed in the pre/maintenance phase. In the variables "Rating of memory" and "Circular recall", although no statistically significant differences were found, we did observe a tendency of

Figure 1
Graphs at the three measurement points. Group 1 is the Experimental Group and 2 is the Control Group. Order of the graphs is Total Score in Metamemory, Organized List, Circular Recall, Preparation of the Object, Associated Pairs and Rating of Memory



the experimental group to improve more than the control group (Figure 1).

DISCUSSION AND CONCLUSIONS

The data from the first assessment indicate the low level of knowledge in people with mental retardation about the processes underlying a memory task, data which confirm our first hypothesis. Moreover, these results concur with the findings of other studies on metamemory and mental retardation (Turner et al., 1997; Turner et al., 1996; Bellinger et al., 1995, Borkowski et al., 1988; Kurtz and Borkowski, 1987; Schneider, Körkel and Weinert, 1987; Valkin et al., 1997; Lodico et al., 1983; Brown, 1978; Campione and Brown, 1977).

The analysis of each one of the subtests shows the participants to be unaware that when we have to remember a task we can use external strategies and notes, or that others can serve as external information stores; in accordance with the studies of Kreutzer, Leonard and Flavell (1975) and Ceci, Lea and Ringstrom (1980), our population functions like pre-school children. They are also unaware that semantic organization favours recall and that few words are recalled better than many, and they do not detect the fact that words with a high degree of association are remembered better than non-associated words. With regard to the variable “Rating of memory”, they tend to underestimate it; this finding does not coincide with that of Brown et al. (1977), who found that that people with mental retardation tended to overestimate their memory capacity.

With regard to the way our participants approached the task, we observed that they were unsure of what to do, and did not ask questions to clarify. Also, they failed to analyze the information presented, and we observed no checking strategy. It can be said that they do not present an ability to plan, understood as the construction and/or use of hierarchical anticipatory representations (plans) for guiding activity (Hoc, 1987).

After the training, the experimental group improved more than the control group in knowledge about their memory, and these data indicate that people with mental retardation can benefit from training in metamemory, which confirms our second hypothesis. What is most remarkable in our view, however, is that they are capable of maintaining this knowledge and even continuing to improve after the training has finished. The participants have learned that organized material can be learned more easily than non-organized material, and that content with a high degree of association is remembered better than non-associated content. This is encouraging with a view to pedagogical intervention, as it appears the adolescents in our study have learned that certain procedures and strategies will help them to improve their recall. In accordance with Biggs (1988), we infer that they are aware of their intentions and motives, of their cognitive abilities and the task demands, and that they are capable of controlling their cognitive resources and their performance.

By way of summary we can state that although people with mental retardation possess scarce knowledge about their memory – metamemory –, they are capable of benefiting from metacognitive training, being able to maintain their improvements after the training. The most significant improvements occurred in the variables of task and strategy. With regard to the person variable Rating of memory, although the results are not statistically significant, we did observe that the experimental group is more realistic, with the disappearance of certain features of helplessness that were noticed in the pre-treatment phase. As far as strategies of control are concerned, they appear in the experimental group in both the post-treatment and maintenance phases.

In future research it will be necessary to determine whether these improvements in metamemory have an influence on memory performance and, if so, on what specific types of memory. In other words, to determine whether this type of population is capable of updating this knowledge, or whether, as Flavell (1978) argues, they will present an executive deficit.

REFERENCES

Belmont, J.M and Borkowski, J.G (1988). A group test of children’s metamemory. Bulletin of the Psychonomic Society. 26, 206-208

Bellinger, E; Borkowski, J.G; Turner, L.A; and Hale C.A. (1995). Perceived task difficulty and intelligence: Determinants of strategy use and recall. Intelligence. 20, 125-143.

Biggs, J. (1988). Approaches to learning and to essay writing. In R. Schmeck (Comp.). Learning strategies and learning styles. N.Y. Plenum Press. 185-228

Borkowski, J.G. (1992). Metacognitive Theory: A framework for teaching literacy, writing and math skills. Journal of Learning Disabilities. 25, 4. 254-257.

Borkowski, J.G; Millstead, M.; and Hale, C. (1988). Components of children’s metamemory: Implications for strategy generalization. In F. Weirnt and M. Permultter (Eds.). Memory development individual differences and universal changes. Hillsdale: LEA.

Borkowski, J.G; Reid, M.K; and Kurtz, B.E. (1984).

Metacognition and retardation: Paradigmatic, theoretical and applied perspectives. In P.H. Broosk, R. Sperber and C. McCauley (Eds.) Learning and cognition in the mentally retarded (pp. 55-75) Hillsdale, N.J.: Erlbaum.

Borkowski, J.G.; and Wanschura, P.B. (1974). Mediational processes in the retarded. In Elli, N.R. (dir.). International review of research in Mental Retardation. Academic Press, N.Y. 7, 1-54.

Brown, A.L. (1978). Knowing when, where and how to remember: a problem of metacognition. In Glaser R. (Ed). Advances in Institutional Psychology. Vol. 1. Laurence Erlbaum Associates, Publishers, Hillsdale, N.J.

Brown, A. and Palincsar, A.S. (1982). Inducing strategic learning from tests by means of informed, self-control training. Topics in Learning and Learning Disabilities. 2, 1-16.

Butterfield, E.C. and Belmont, J.M. (1977). Assessing and improving the retarded. In N.R. Ellis (Ed). International review of research in mental retardation. Vol. 7, N.Y. Academic.

Campione, J.C. and Brown, A.L. (1977). Memory and metamemory development in educable retarded children. In R. Kail and J. Hagen (Eds.). Metacognition, motivation and understanding. Hillsdale: LEA.

Campione, J.C.; Brown, A.L.; and Ferrara, R. (1982). Mental retardation and intelligence. In Sternberg, R. (Ed.). Handbook of Human Intelligence. Cambridge University Press. N.Y. (Spanish translation: Inteligencia Humana. Vol.2. Paidós (1987).

Ceci, S.J.; Lea, S.E.; and Ringstrom, M.D. (1980). Coding processes in normal and learning-disabled children: Evidence for modality-specific pathways to the cognitive system. Journal of Experimental Psychology: Human Learning and Memory. 6. 785-797.

Cox, B.D. (1994). Children's use of Mnemonic Strategies: Variability in Response to Metamemory Training. The Journal of Genetic Psychology. 155 (4), 423-442.

Flavell, J.H. (1978). Metacognitive development. In Scandura, J.M and Brainerd, C.J. (Eds.). Structural process theories of complex human behaviours. Alphen A.D. Rija: Siggthoff and Noordehoff.

Flavell, J.H. (1987). Speculations about the nature and development of metacognition. In F. Weinert and R. Kluwel (Eds.). Metacognition, motivation and undertanding. Hillsdale: LEA.

Flavell, J.H. and Wellman, H.M. (1977). Metamemory. In R.V. Kail and J.W. Hagen (Eds.). Perspectives on the development of memory and cognition. Hillsdale, LEA.

García, E. (1997). Inteligencia y Metaconducta. Revista de Psicología General y Aplicada. 50, 297-312.

Hoc, J.M. (1987). Psychologie cognitive de la planification. Grenoble: PUG.

Kreutzer, M.A.; Leonard, C.; and Flavell, J.H. (1975). "An interview study of children's knowledge about memory". Monographs of the Society for Research in Child Development, 40. (1, nº 159). pp. 1-58.

Kurtz, B.E. and Borkowski, J.G. (1987). Development of strategic skill in impulsive and reflective children: A longitudinal study of metacognition. Journal of Experimental Child Psychology. 43. 129-148.

Lodico, M.G.; Ghalata, E.S.; Levin, J.R.; Pressley, M.; and Bell, J.A. (1983). The effects of strategy monitoring training on children's selection of effective memory strategies. Journal of Experimental Child Psychology, 35, 263-277

Mayor, J. (1988). Manual de Educación Especial. Madrid. Anaya.

Mayor, J. and Saínz, J. (1988). Procesos cognitivos en la deficiencia mental. In J. Mayor (Ed.). La Psicología en la Escuela Infantil. Madrid. Anaya.

Osman, M.E. and Hannafin, M.J. (1992). Metacognition research and theory: Analysis end implications for instructional design. Educational Technology Research and Development. 40, 2. 83-99.

Schneider, W.; Körkel, J.; and Weinert, F.F. (1987). The effects of intelligence, self-concept, and attributional style on metamemory and memory behaviour. International Journal of Behavioral Development. 281-299.

Sternberg, R. (1985). Beyond IQ: A triarchic theory of human intelligence. Cambridge. Cambridge University Press.

Turner, L.; Hale, C.; and Borkowski, J.G. (1996). Influence of Intelligence on Memory Development. American Journal on Mental Retardation. Vol. 100. nº. 5. 468-480.

Turner, L.; Hale, C.; Wilcox, M.; and Borkowski, J.G. (1997). The emergence of strategies in EMR and nonretarded children: Influences from home. Paper presented at the Gatlinburg conference on research and theory on mental retardation and developmental disabilities. Gatlinburg, TN.

Valkil, E.; Shelef-Reshef, E.; and Levy-Shiff, R. (1997). Procedural and declarative memory processes: Individuals with and without mental retardation. American Journal on Mental Retardation. Vol., 102, nº2, 147-160.