6

The Us as a Couple Questionnaire (US): A Unidimensional Measure of Couple Satisfaction

Cor de Jong, Laura DeFuentes-Merillas, Florieke van de Werken, and John Birtchnell

Introduction

Relationship quality and satisfaction are two of the most widely investigated areas of relationship research (Boesch et al. 2007; Karney and Bradbury 1995; Rosen-Grandon et al. 2004) which tend to stem from two different approaches: the unidimensional and multidimensional approach (Kluwer 2001).

Crosby (1991) considers relationship quality a unidimensional concept and defines it as 'the global evaluation of marriage, i.e. the evaluation placed on the relationship as a whole by the marital partners' (p. 3). Others consider it as a multidimensional higher-order construct. For instance, the Perceived Relationship Quality Components (PRQC) Inventory (Fletcher et al. 2000) discriminates five subscales: relationship satisfaction, commitment, intimacy, trust, passion, and love.

The PRQC inventory conceptualises relationship satisfaction as part of relationship quality as it is by others (e.g. Hassebrauck and Aron

C.I. de Jong (✉) • L. DeFuentes-Merillas • F. van de Werken • J. Birtchnell
UK

2001; Hendrick et al. 1998). A commonly used definition of relationship satisfaction is 'positivity of affect or attraction to one's relationship' (Rusbult 1983, p. 102). Additional measures that assess relationship satisfaction are: the Marital Satisfaction Inventory (Snyder 1979); the Relationship Assessment Scale (Birnbaum 2007; Hendrick et al. 1998); the Relationship Satisfaction Scale (Heyman et al. 1994; Troy et al. 2006); the Revised Dyadic Adjustment Scale (Busby et al. 1995), and the RELATionship Evaluation (Busby et al. 2001; Larson et al. 2007). Some measures are quite lengthy and measure various dimensions, like the 280-item Marital Satisfaction Inventory (Snyder 1979) or the 271-item RELATionship Evaluation (Busby et al. 2001) which measures the influence of individual, family, and cultural contexts on relationship satisfaction.

Most questionnaires have been developed according to a classical test model, which assigns a measure on a scale as the sum of the responses to each item on that scale (Crocker and Algina 1986; Nunnally and Bernstein 1994). The Rasch model (Robin et al. 1999) is another approach which constructs a hypothetical unidimensional line along which items and persons are located according to their difficulty and ability to measure underlying concepts. This provides a scale on which the items are placed hierarchically: the simplest items will be answered by all respondents, and the more difficult items only by those who are best informed about the concept. This is in contrast to most questionnaires, for which items of a subscale are related but not hierarchically. So in these questionnaires it is not clear from a subscale score which items are summed. Fit statistics such as R1 and Q2 (see Suarez-Falcon and Glas 2003 for further explanation of the Rasch statistics) indicate how well different items describe the group of respondents and how well individual respondents fit the group (Wright and Masters 1982; Wright and Stone 1979).

A practical advantage of Rasch homogeneous, unidimensional questionnaires is their use in clinical settings. When a couple with relationship problems is in therapy, it can be useful to gauge their general evaluation of the relationship from the start. For a good relationship partners do not have to think the same things, but if they are not good friends any more one can wonder whether the relationship is still worth it. On a Rasch homogeneous scale a decrease of the scale score over time indicates that

more difficult aspects of the concept measured have been solved. In other words: the higher the score, the worse the relationship. But it is not just the score that makes Rasch analysis interesting for testing the psychometric quality of questionnaires. The sequence revealed by the analysis makes interpretation easier and is appealing for a clinician. If for instance an item halfway through the questionnaire is positive then lower items will also be positive.

Unidimensional measures of relationship satisfaction measure the relationship in general (Hendrick et al. 1998; Norton 1983) and the results are usually unambiguous (Fincham and Linfield 1997), meaning that the most difficult issues are solved. These instruments usually consist of just a few (less than ten) items, with questions like 'In general, how satisfied are you with your relationship?' For example, the Interactional Problem Solving Questionnaire (Lange 1983) contains only four questions on general satisfaction.

The Us as a Couple Questionnaire (US; Birtchnell 1999) is a brief, single-scale 20-item measure which assesses how well each partner considers the two partners get on together. The US shows similarities with the Family Adaptability and Cohesion Evaluation Scales (FACES III; Olson 1986) and Marital Stability and Cohesion Evaluation Scales (MACES III, Olson 2000) which are used predominantly with couples. The FACES III is a 20-item self-report questionnaire that measures the amount of cohesion and adaptation within a relationship according to the partners. It is administered twice per participant: once in terms of the way one currently experiences the relationship, and once in terms of the way this relationship would ideally be. The difference between these two measures provides an indication of the family or couple satisfaction: the less the difference, the more satisfied the partners are (Olson 2000). Unlike the FACES III, the US evaluates the relationship on one dimension by asking how satisfied each partner is with his/her relationship. Compared to the FACES III, the US has a narrower focus and the items are constructed from a 'we' perspective (e.g. 'We found ourselves avoiding each other' and 'There is a lot of give and take in our relationship').

Before the development of the US, Birtchnell developed the Couples Relating to Each Other Questionnaire (CREOQ; Birtchnell et al. 2006)

which measures negative forms of interrelating within couples and consists of a set of four questionnaires of 48 items each. The US was developed as an additional, less complex instrument that could provide a global indication of how a partner judges the aspect of getting on together in his/her relationship. The US may assist couples in clarifying perceptions and highlight areas of agreement and disagreement within the relationship. Further, it shows the individual's and the couples' strengths and challenges related to the quality of their relationship. The self-report nature of the US means that it can be easily administered in research and therapy settings. However, the US has not yet been validated psychometrically, particularly in terms of its unidimensionality, and this was the aim of the current study. Five hypotheses were tested: (1) the items of the US will fit the Rasch model; (2) the internal reliability of the US is sufficient; (3) the average total score of the US will be higher for 'addicted couples' than for 'non-addicted couples'; (4) there will be a significant negative correlation between the US and the Satisfaction subscale of the Interactional Problem Solving Questionnaire, and (5) there will be no significant correlation between the US and the Symptoms Checklist-90-Revised (SCL-90-R).

Method

Participants

Two groups participated in this study: 112 Dutch couples that reported no psychological or relationship problems (Group 1) and 50 couples, in which one of the partners had an alcohol or substance use disorder for which they were being treated in an addiction treatment centre (Group 2). All participants were heterosexual.

Group 1 was recruited from the general Dutch population via snowball sampling (Biernacki and Waldorf 1981). The third author asked acquaintances to invite their own acquaintances to participate in a study about the Dutch version of the US. Those willing to participate were asked to contact the researcher. To be eligible, no partner in a couple could be

suffering from any substance abuse or psychological disorder. The mean age of Group 1 was 33.1 years (SD = 12.8) and the mean duration of their relationship was 9.9 years (SD = 9.3).

Group 2 consisted of 50 couples, of which one of the partners was in treatment for alcohol or drug dependence (80 % alcohol, 2 % benzodiazepines, 2 % opioids, or 16 % more than one substance). They were recruited by therapists working at one of three addiction treatment centres in the Netherlands. Couples were included if the partner was not dependent on any psychoactive substances according to DSM-IV criteria. The patients were following a period of three-month inpatient treatment. The mean age of Group 2 was 45 years (SD = 8.5), the mean duration of the relationship was 19 years (SD = 10.6), and the mean duration of addiction problems was 11 years (SD = 8.3).

Measures

The Us as a Couple Questionnaire (US; Birtchnell 1999)

The US measures how each partner considers the two partners get on together. It has 20 items (Table 6.1), each with a possible response of 'true' or 'false'. Each even item answered with a 'true' response and each odd item answered with a 'false' response receives one point. The higher the total score, the less satisfied one is with his/her relationship. The minimum possible total score is 0 and the maximum is 20.

Interactional Problem Solving Questionnaire (IPSQ; Lange 1983)

The IPSQ contains four questions on global satisfaction about relationships. The higher the total score, the more satisfied one is with his/her relationship. This is contrary to the US, where a high score indicates difficulties in the relationship. The IPSQ has a good level of internal reliability (Cronbach's alpha = 0.81).

Table 6.1 The original English and translated Dutch items of the US and the Rasch order of the 17 items of the US-2

Original English items	Items translated in Dutch	Rasch order
1. We are good friends. ($F = 1$)	1. We zijn goede vrienden.	17.
2. We find ourselves avoiding each other. ($T = 1$)	2. We komen erachter dat we elkaar vermijden.	13.
3. We help each other out of difficulties.* ($F = 1$)	3. We helpen elkaar uit de moeilijkheden.*	* Deleted from original US
4. We don't have very much in common. ($T = 1$)	4. We hebben niet erg veel gemeenschappelijk.	5.
5. When we have a problem we sit down and sort it out together. ($F = 1$)	5. Als we een probleem hebben, gaan we bij elkaar zitten en zoeken we het samen uit.	4.
6. We seem to be drifting apart. ($T = 1$)	6. We lijken van elkaar te vervreemden.	8.
7. There is a lot of give and take in our relationship. ($F = 1$)	7. Er is een goede balans tussen geven en nemen in onze relatie.	3.
8. We can't talk for long without starting to argue. ($T = 1$)	8. We kunnen niet lang met elkaar praten zonder dat we gaan redetwisten.	6.
9. If we have a row it is quickly over and there are no hard feelings afterwards. ($F = 1$)	9. Als we ruzie hebben is het snel over en zijn er geen verwijten achteraf.	2.
10. The fun has gone out of our relationship. ($T = 1$)	10. De lol is er af in onze relatie.	14.
11. We are usually able to see each other's point of view. ($F = 1$)	11. We zijn meestal in staat elkaars standpunt in te zien.	11.
12. We rub each other up the wrong way. ($T = 1$)	12. We strijken elkaar meestal tegen de haren in.	12.
13. We are quite open with each other. ($F = 1$)	13. We zijn behoorlijk open naar elkaar.	7.
14. We don't enjoy each other's company.* ($T = 1$)	14. We genieten niet van elkaars gezelschap.*	*Deleted from original US
15. When we each want different things, we compromise.* ($F = 1$)	15. Als we allebei iets anders willen, komen we tot overeenstemming.*	*Deleted from original US
16. We don't have much to say to each other. ($T = 1$)	16. We hebben elkaar niet veel te zeggen.	10.

(*continued*)

Table 6.1 (continued)

Original English items	Items translated in Dutch	Rasch order
17. We often find ourselves thinking the same thing. ($F = 1$)	17. We komen erachter dat we hetzelfde denken.	1.
18. We are not good for each other. ($T = 1$)	18. We zijn niet goed voor elkaar.	15.
19. We are always pleased to see each other. ($F = 1$)	19. Het doet ons altijd plezier om elkaar te zien.	9.
20. We can't seem to agree about anything. ($T = 1$)	20. Het lijkt erop dat we het nergens over eens zijn.	16.

Note: All even items are scored with one point if true and all odd items are scored with one point if false

Symptoms Checklist-90-R (SCL-90-R; Arrindell and Ettema 1986)

The 90-item SCL-90-R measures physical and psychological complaints during the past week (somatisation, obsessive-compulsiveness, inter-personal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, and psychoticism). The Dutch version of the SCL-90-R has good psychometric properties (Arrindell and Ettema 1986). It was administered to Group 1.

Dutch Translation of the US (Hunt et al. 1991)

Two Dutch researchers translated the English version of the US to Dutch independently. Another independent researcher, whose native language was English and whose second language was Dutch, translated the Dutch version of the questionnaire back into English. The original and back-translated versions were compared and differences were discussed. Eventually the Dutch version was established. Comments on the questionnaire were gathered from three therapists from the family therapy in addiction treatment centres. A pilot study on five patients showed that the questionnaire could be easily read and the questions were well understood.

Procedure

Couples in both groups provided informed consent and were administered the US and the Satisfaction subscale of the IPSQ. Group 1 also completed the SCL-90-R and questions about psychiatric and substance-related disorders. The partners were instructed not to view each other's responses. Scores were calculated by computer. The medical ethical board of the addiction treatment centres approved the procedure.

Data Analysis

The structure of the US was analysed in order to test whether the items fit a Rasch model and whether it could be regarded as unidimensional. In cases where the items did not fit, items were deleted using the Rasch Scaling Program (Robin et al. 1999). The internal consistency of the US was tested by computing the Cronbach alpha coefficient. The comparison of the US total score between the two groups was tested with an independent-samples t-test and the correlations between the US and the Satisfaction subscale of the IPSQ and between the US and the SCL-90-R were estimated using Pearson's correlation coefficient.

Results

Unidimensionality of the US

A Rasch analysis was conducted on the two combined groups described above which was divided into the scores of two gender groups. Table 6.2 shows that the Rasch model did not fit the US data for either group. After deleting the three items which contributed most to the significance of R1 and Q2, the tests for both gender groups became insignificant. This indicated that the remaining set of items was unidimensional. Thus, the Rasch model for the remaining 17 items fit the data for both males and females (see Table 6.2). The deleted items were: 3: *We help each other out of difficulties*, 14: *We don't enjoy each other's company*, and 15: *When we*

Table 6.2 Results of the Rasch analysis for the original 20-item US and 17-item US-2

	R1 statistic	Df	*p* (R1)	Q2 statistic	Df	*p* (Q2)
Original version (20 items)						
Males	61.57	38	0.01	488.95	510	0.74
Females	49.79	38	0.10	663.61	510	0.00
US-2 (17 items)						
Males	44.60	32	0.07	288.62	357	1.00
Females	44.08	32	0.08	312.77	357	0.96

each want different things, we compromise. Items 3, 14 and 15 were therefore excluded from the subsequent analyses. This short version of the US is hereafter referred to as the US-2.

The sequence of US-2 items based on the Rasch analysis is listed in the last column of Table 6.1. The item '*We often find ourselves thinking the same thing*' (item 17) was the one that most couples disagreed on. The item '*We are good friends*', was the one that most couples agreed on, that is the majority of couples reported that they were good friends. Furthermore, the results of the Rasch analysis indicated that the US-2 is unidimensional. This can be interpreted as follows. If a partner agrees on the statement '*We seem to be drifting apart*' (item 6, sequence order 8) then it is highly probable that he/she will also agree on the item '*We don't have very much in common*' (item 4, sequence order 5) and that he/she will disagree on the item '*There is a lot of give and take in our relationship*' (item 7, sequence order 3).

Internal Consistency

The internal consistency of the 17 items of the US-2 was very good for both males and females (Cronbach's alpha = 0.79 and 0.81, respectively).

Discriminant Validity

It was expected that the US mean scores would differ for the two groups. The means and standard deviations for Groups 1 and 2 were 1.44 (1.79) and 4.71 (3.71), respectively and the difference between the two groups

was statistically significant ($t = -8.45$; $p < 0.001$; 95 % CI [–4.04, –2.50]. Therefore, the Dutch version of the US clearly discriminates community control couples from couples in treatment for an addiction.

Construct Validity

Correlations between the US-2, the Satisfaction subscale of the IPSQ and the total score of the SCL-90-R are presented in Table 6.3. As predicted, there was a statistically significant negative correlation between the US-2 and the Satisfaction subscale for both groups. The correlations between the US-2 and the SCL-90-R were all weak and non-significant.

Discussion

The findings of this study indicate that the revised version of the US (the US-2) which consists of 17 of the original 20 items fit well with the Rasch model, meaning that the scale is unidimensional. This yielded substantial support for the construct validity of the US-2. Due to the scale's unidimensionality, the sum of all items assesses the level of relationship satisfaction. Because the items are hierarchically ordered from a high level of satisfaction (low score on the US-2) to a low level (high score on the US-2), a decrease of the score is easily interpreted. It also means that if two individuals have the same total score, they have answered the same questions positively. One of the advantages of a measure fitting the

Table 6.3 Correlations between the US-2 and Satisfaction subscale of the IPSQ and between the US-2 and SCL-90-R for males and females

	US-2 – IPSQ		US-2 – SCL-90-R	
Group	Male	Female	Male	Female
1	−0.516**	−0.708**	0.084	0.171
2	−0.577**	−0.608**	×	×

Group 1 = community-based population ($n = 112$)
Group 2 = couples with one addicted partner ($n = 50$)
*$p < 0.05$; **$p < 0.01$; × No correlation calculated because the SCL-90 was only administered to sample 1

Rasch model is that one group is well comparable with another group. Furthermore, the Cronbach alphas ranged from 0.79 to 0.88, showing that the US-2 has good reliability.

Couples in which a partner had an addiction (Group 2) scored significantly higher on the US-2 than the community group (Group 1). As expected, the addiction problem of one partner had a negative effect on the quality of his/her relationship; in our study couples in which a partner had an addiction had a lower mean score of 4.2 (identical for both males and females) than couples seeking therapy in Birtchnell and Spicer's unpublished study (8.8 for males and 10.5 for females), which indicates that addicted couples tended to report less relationship satisfaction than non-addicted couples. It should be noted, however, that the mean total score of the therapy couples in Birtchnell and Spicer's study (US: 20 items) was three points higher than the mean total score of the addicted couples (US-2: 17 items).

The current results indicate that the US-2 has a good level of convergent validity. As predicted, the US-2 was negatively related to the Satisfaction subscale of the IPSQ which suggests that the US-2 is a valid instrument for assessing aspects of satisfaction in relationships. The findings also indicated an acceptable level of divergent validity since all correlation coefficients between the US-2 and the SCL-90-R were weak and non-significant.

Our study has several limitations. Firstly, it was conducted in the Netherlands involving couples in which one partner was being treated for a substance-related disorder and so there may be different results in populations in which one partner has a different mental disorder. In addition, because of the rather small samples it was not possible to take into account potential confounding couple-related variables such as length of the relationship and the presence of children in the household.

Overall, the results indicated that the US-2 is a unidimensional instrument fitting the Rasch model with good internal consistency, validity, and reliability. It is a short self-report measure which is straightforward to complete and does not require much training for the professionals (e.g. therapists) to administer it. This makes the US-2 a relatively inexpensive research and treatment evaluation instrument. The US-2 is designed to be used as a global measure of how both partners consider they get on

together; for different aspects of relationship difficulties, such as communication or sexuality, other questionnaires are required. Because both partners complete the US-2, the evaluation of the relationship is more objective than if only one person were to judge his/her relationship. We expect the US-2 to measure change in a relationship, but this has only be assessed in a pilot study (DeJong et al. 2008). The relationship of the US-2 with the concepts 'relationship quality' and 'relationship satisfaction' is also not yet clear and so this would be a fruitful area for future research.

References

Arrindell, W. A., & Ettema, J. H. M. (1986). SCL-90: Manual for a multidimensional indicator for psychopathology [SCL 90: Handleiding bij een multidimensionele psychopathologie-indicator]. Lisse: Swets Zeitlinger

Biernacki, P., & Waldorf, D. (1981). Snowball sampling: Problems and techniques of chain referral sampling. *Sociological Methods and Research, 10*(2), 141–163.

Birnbaum, G. E. (2007). Attachment orientations, sexual functioning, and relationship satisfaction in a community group of women. *Journal of Social and Personal Relationships, 24*(1), 21–35.

Birtchnell, J. (1999). *Relating in psychotherapy: The application of a new theory.* London: Praeger.

Birtchnell, J., Voortman, S., DeJong, C. A. J., & Gordon, D. (2006). Measuring interrelating within couples: The Couple's Relating to Each Other Questionnaires (CREOQ). *Psychology and Psychotherapy: Theory, Research and Practice, 79*(3), 339–364.

Birtchnell, J., & Spicer, C. (unpublished). A new interpersonal system for describing and measuring the relating of marital partners. Available from the first author; Department of Psychiatry, Institute of Psychiatry, London.

Boesch, R. P., Cerqueira, R., Safer, M. A., & Wright, T. L. (2007). Relationship satisfaction and commitment in long-term male couples: Individual and dyadic effects. *Journal of Social and Personal Relationships, 24*(6), 837–853.

Busby, D. M., Christensen, C., Crane, D. R., & Larson, D. H. (1995). A revision of the Dyadic Adjustment Scale for use with distressed and nondistressed couples: Construct hierarchy and multidimensional scales. *Journal of Marital and Family Therapy, 21*(3), 289–308.

Busby, D. M., Holman, T. B., & Taniguchi, N. (2001). RELATE: Relationship evaluation of the individual, family, cultural, and couple contexts. *Family Relations, 50*(4), 308–317.

Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando, FL: Holt, Rinehart, & Winston.

Crosby, J. F. (1991). Cybernetics of cybernetics in assessment of marital quality. *Contemporary Family Therapy, 13*(1), 3–15.

DeJong, C. A. J., Mellink, D. C., & DeJong-Verhagen, J. G. (2008). A short interpersonal intervention is dyades: A pilot-study. [Een korte interpersoonlijke interventie bij dyades: een pilot-study]. *Systeemtherapie, 20*(2), 76–89.

Fincham, F. D., & Linfield, J. L. (1997). A new look at marital quality: Can spouses feel positive and negative about their marriage? *Journal of Family Psychology, 11*(4), 489–502.

Fletcher, G. J., Simpson, J. A., & Thomas, G. (2000). Ideals, perceptions, and evaluations in early relationship development. *Journal of Personality and Social Psychology, 79*(6), 933–940.

Hassebrauck, M., & Aron, A. (2001). Prototype matching in close relationships. *Personality Social Psychology Bulletin, 27*(9), 1111–1122.

Hendrick, S. S., Dicke, A., & Hendrick, C. (1998). The Relationship Assessment Scale. *Journal of Social and Personal Relationships, 15*(1), 137–142.

Heyman, R. E., Sayers, S. L., & Bellack, A. S. (1994). Global marital satisfaction versus marital adjustment: An empirical comparison of three measures. *Journal of Family Psychology, 8*(4), 432–446.

Hunt, S. M., Alonso, J., Bucquet, D., Niero, M., Wiklund, I., & McKenna, S. (1991). Cross-cultural adaptation of health measures. European Group for Health Management and Quality of Life Assessment. Review. *Health Policy, 19*(1), 33–44.

Karney, B. R., & Bradbury, T. N. (1995). The longitudinal course of marital quality and stability: A review of theory, method, and research. *Psychological Bulletin, 118*(1), 3–34.

Kluwer, E. S. (2001). The quality of intimate relationships: An overview [De kwaliteit van intieme relaties: een overzicht]. *Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden, 56*, 138–152.

Lange, A. (1983). *Interactional Problem Solving Questionnaire [Interactionele Oplossings Vragenlijst]. Manual and Research [Handleiding en onderzoek]*. Deventer: Van Loghum Slaterus.

Larson, J. H., Vatter, R. S., Galbraith, R. C., Holman, T. B., & Stahmann, R. F. (2007). The RELATionship Evaluation (RELATE) with therapist-assisted interpretation: Short-term effects on premarital relationships. *Journal of Marital Family Therapy, 33*(3), 364–374.

Norton, R. (1983). Measuring marital quality: A critical look at the dependent variable. *Journal of Marriage and Family, 45*(1), 141–151.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NJ: McGraw-Hill.

Olson, D. H. (1986). Circumplex model VII: Validation studies and FACES III. *Family Process, 25*, 337–351.

Olson, D. H. (2000). Circumplex model of marital and family sytems. *Journal of Family Therapy, 22*(2), 144–167.

Robin, F., Xing, D. H., & Hambleton, R. K. (1999). Rasch Scaling Program. *Applied Psychological Measurement, 23*(1), 90–94.

Rosen-Grandon, J. R., Myers, J. E., & Hattie, J. A. (2004). The relationship between marital characteristics, marital interaction processes, and marital satisfaction. *Journal of Counseling and Development, 82*(1), 58–68.

Rusbult, C. E. (1983). A longitudinal test of the investment model: The development (and deterioration) of satisfaction and commitment in heterosexual involvements. *Journal of Personality and Social Psychology, 45*(1), 101–117.

Snyder, D. K. (1979). Multidimensional assessment of marital satisfaction. *Journal of Marriage and Family, 41*(4), 813–823.

Suarez-Falcon, J. C., & Glas, C. A. (2003). Evaluation of global testing procedures for item fit to the Rasch model. *British Journal of Mathematical and Statistical Psychology, 56*, 127–143.

Troy, A. B., Lewis-Smith, J., & Laurenceau, J. P. (2006). Interracial and intraracial romantic relationships: The search for differences in satisfaction, conflict, and attachment style. *Journal of Social and Personal Relationships, 23*(1), 65–80.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago, IL: MESA Press.

Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago, IL: MESA Press.