The Journal of Parapsychology, Vol. 61, September 1997

A META-ANALYTIC COMPARISON OF THE SENSITIVITY OF DIRECT HITS AND SUMS OF RANKS AS OUTCOME MEASURES FOR FREE-RESPONSE STUDIES

BY JULIE MILTON

ABSTRACT: For years, researchers have been applying a variety of statistics to obtain outcome measures for free-response extrasensory perception studies without knowing which measure is the most sensitive. Recently the choice has tended to be between direct hits and sums of ranks, with direct hits the most popular. However, according to different theoretical models of psi functioning and of judging practices, either measure could be superior. If researchers are using the wrong measure then statistical power is lost and more (time-consuming) trials than necessary must be conducted in a study to obtain either statistical significance or an accurate estimate of effect size. In order to determine whether direct hits or ranks are the more sensitive measure, I sought both outcome measures for all free-response ESP studies published in the last five years. Data were obtained for 46 studies, including 42 ganzfeld studies. Overall, the cumulation of the studies' outcomes was statistically significantly above chance. On average, sums of ranks outperformed direct hits in terms of both effect sizes and *p* values, but not to a statistically significant degree.

In a typical open-deck free-response extrasensory perception study, a target is randomly selected from a set of four pictures. The picture is kept hidden from the receiver, who then attempts to describe its content. The receiver is then shown a duplicate set of the four pictures (to avoid handling cues). Still not knowing which is the target, the receiver (or an independent judge who is blind to the target's identity) rates or ranks the four pictures in terms of how well each corresponds to the earlier attempted description. This procedure is repeated on other trials until the study is completed.

In recent years, two measures—direct hits and ranks—have become the most popular ways of representing such a study's outcome.[1] A trial

I am grateful to Daryl Bem, Dick Bierman, Richard Broughton, Jim Carpenter, Kathy Dalton, H. Kanthamani, Tony Lawrence, Bruce McDonough, John Palmer, Adrian Parker, James Spottiswoode, Caroline Watt, Rens Wezelman, Melvyn Willin, and Nancy Zingrone for generously taking the time and trouble to provide me with unpublished data from their studies for this meta-analysis.

[1] Parapsychologists have used two other outcome measures for this type of study. Standardized ratings (Stanford & Mayer, 1974) are now rarely used as a study's main outcome measure because of concerns that their distribution is likely to be far from normal (Hansen, 1986; Schouten & Camfferman, 1988). Binary hits have almost disappeared from free-response research, probably because of their likely insensitivity.

obtains a direct hit if the target picture obtains the highest rating or rank, and otherwise is counted a miss. The study's outcome is assessed by means of the binomial distribution, with $p = 1/4$. If, instead, ranks are being used as the study's outcome measure, then the rank assigned to the target is the trial's outcome and the study's outcome can be determined by comparing the sum of ranks across all the trials to the chance expected distribution (Solfvin, Kelly, & Burdick, 1978).

Although researchers, in effect, find themselves faced with a choice between direct hits and sums of ranks, there are no clear indications at present as to which—if either—of the two is the more sensitive measure. However, it is important that if one measure is superior to another then that measure should be used because effect sizes in free-response ESP studies are very low. In 1986, Utts used the then current estimate of ganzfeld ESP study effect size of a hit rate of 33% (MCE being 25%) to calculate that over 100 trials would be needed to give a ganzfeld study a 50:50 chance of reaching statistical significance (Utts, 1986). Recent research indicates that ganzfeld studies conducted after Utts' calculations are not achieving even that effect size (Milton & Wiseman, 1997). It clearly makes sense for researchers to be sure that they are using the most sensitive measure of psi available in order to achieve the maximum statistical power with the minimum of time-consuming and expensive trials.

As far as theoretical arguments about which measure should be better are concerned, no clear winner emerges. Ranks might be expected to be the more sensitive measure of ESP performance because they use more information from the trial than do judgements of direct hits, which do not distinguish between near misses and clear failures. However, as Hansen and Utts (1987) point out, whether sums of ranks are in practice more sensitive than direct hits depends both upon how ESP manifests itself and on how judges behave when assigning ranks. If psi operates in an all-or-nothing fashion in each trial, so that it either operates so strongly that a direct hit is certain, or does not operate at all so that each rank is equally likely, then direct hits, not ranks, might be the most sensitive measure. If, instead, psi operates like a weak signal hidden in noise, then the frequency distribution of ranks should tend to shift toward the lower-rank, above-chance end and ranks will be the more sensitive measure. However, in either case, the judge's behavior can affect which measure should be best. Even if psi acts in an all-or-nothing manner, a judge who finds it hard to discriminate which judging-set item best matches the receiver's response might often turn a direct hit into a second, or, less frequently, even a third rank, in which case ranks would once again be the more sensitive measure. This situation could easily arise if judging sets contain items that are not sufficiently different from each other. Conversely, even if psi acts as a signal buried in noise, a judge who pays attention only to accurately assigning a rank of 1 to the item that best matches the response and ignores the other judging-set items, assigning their ranks more or less at random, will make direct hits the more sensitive measure.

The empirical data examined so far shed no more light upon the question than the theoretical arguments. Only Honorton (1995) has examined data from real studies to attempt to determine which measure might be best, in a database of 73 ganzfeld studies. Although he states, "It is clear from the meta-analysis that the various alternatives to direct hits assessment have not been effective" (p. 134), he reports no details of what data were compared or of any statistical analyses. It may be that very few studies reporting rank data were included in the comparison. Whether or not this was the case, it appears that the comparison was made between studies, which makes it possible that any relative superiority of direct hits over other measures may not have been due to differences in the sensitivity of the outcome measures but to other factors in the studies.

The most direct way to address the question is to perform a meta-analysis of within-study comparisons of ranks and direct hits. The present author therefore set out to obtain both direct hit and rank data from a database of recent free-response studies and compared the success of the two measures. It was decided in advance to conduct the planned analyses both on the whole database and separately for the subset of ganzfeld studies. Ganzfeld studies were examined separately because they constitute a research domain of special interest. Moreover, features associated with the ganzfeld, such as slight possible disorientation of the receiver immediately after the response period and during the judging, might moderate any differences between hit- and rank-based outcome measures, in which case separate examination is again desirable.

Method

Study Retrieval

The author searched the main English language parapsychology journals (*European Journal of Parapsychology, Journal of the American Society for Psychical Research, Journal of Parapsychology, Journal of the Society for Psychical Research)* and the annual Parapsychological Association convention's *Proceedings.* Bem and Honorton's (1994) well-known autoganzfeld paper was also chosen to be included before data collection began.

Because direct hits are now so popular as a main outcome measure, it seemed likely that sums of ranks might be reported in published studies as a secondary matter of interest only when they were successful but not otherwise. Failure to retrieve many unpublished sums of ranks might therefore have made it appear that sums of ranks were more successful than they really were. In order to have a good chance of retrieving all studies within a given publication period, the author chose to examine only studies published in 1992 or later, because it seemed likely that many authors would have kept such recent data, particularly if they were following the recommendation of the American Psychological Association (Publication Manual, 1994) or of the Parapsychological Association (Stanford, 1990) that authors retain for five years data appearing in those organizations' publications.

Inclusion Criteria

Only studies that had the receiver or independent judge rate or rank the target-set items in an open-deck study were included in the database.

The Database

Fifty-four studies, including 44 ganzfeld studies (see Appendix) were retrieved from the literature search. Only 8 of the studies reported both direct hit and rank data but authors of 38 out of the remaining 46 studies responded to a request for the relevant unreported data. Data were therefore available for 46 of the 54 studies of all types (85%) and for 42 out of the 44 retrieved ganzfeld studies (95%). Of the 37 studies that reported either direct hits or sums of ranks as the sole main outcome measure, 33 (89%) used direct hits. Only 3 studies used independent judging as the sole means of judging; in 4 cases both receivers' and independent judges' data were provided and in the remaining studies, receivers (often helped by experimenters) did the judging.

RESULTS[2]

All analyses, including the statistics used, whether tests were one- or two-tailed, and the level of alpha (.05), were prespecified in writing

---

[2]A few data in this paper differ very slightly from those presented in an earlier version presented at the 1997 annual convention of the Parapsychological Association because of the erroneous inclusion in the earlier paper of some judges' preference data (their rankings of the target-set items in order of liking) in addition to correspondence judging data in the rank calculations for two studies.

before data collection for the review began, unless otherwise stated. It was also decided in advance of data collection to treat each whole experiment as the unit of analysis, unless the experiment contained both ganzfeld and non-ganzfeld trials, in which case the two groups of trials were to be treated as separate studies. It was also prespecified that if more than one person's judging data were reported in the paper (such as when both the receiver and an independent judge evaluated the target sets), then all the judges' data would be combined, whenever available, to give a mean number of direct hits and mean sum of ranks for the experiment.

When the average of several judges' ratings gave a number of direct hits or a rank sum that was an integer plus 0.5, the average was rounded up to the next integer in the conventional manner. However, this procedure slightly favors direct hits over ranks. Rounding-up increases the number of hits, leading to a larger value of z, but increases the sum of ranks, which leads to a lower value of z because smaller target ranks are associated with higher performance. There were only four instances in which such rounding-up occurred out of 92 data points but post hoc analyses using rounding-down to the nearest integer are also reported to show any effects of this systematic error.

Overall Cumulation

For each study, the standard normal deviate (z) associated with its number of direct hits[3] and sum of ranks[4] was obtained. For both measures, z was obtained from the upper tail probability associated with the outcome. Table 1 shows the Stouffer zs obtained for direct hits and sums of ranks for both the ganzfeld and whole sample of studies, and for both rounded-up data and rounded-down data. There was statistically significant evidence of an above-chance effect for both measures, both for the whole sample of studies and the ganzfeld subset, regardless of whether data were rounded up or down.

---

[3] Most exact probabilities for sums of ranks were calculated using a computer program written by Paul Stevens (Milton & Stevens, 1997). When the number of trials was too high for computer memory to deal with, I used the normal approximation to the sum of ranks distribution.

[4] I am grateful to Norman S. Don and Bruce E. McDonough for providing me with a copy of their computer program in QBASIC (also available in FORTRAN) for calculating exact binomial probabilities. The program can be obtained from them at the Kairos Foundation, 405 N. Wabash Ave., Suite 2201, Chicago, IL 60611, USA.

TABLE 1
STOUFFER z VALUES FOR DIRECT HITS AND SUMS OF RANKS FOR GANZFELD STUDIES AND ALL STUDIES (WITH ONE-TAILED p VALUES GIVEN IN PARENTHESES)

	Ganzfeld studies (N = 42)		All studies (N = 46)	
	Hits	Ranks	Hits	Ranks
Rounded-up data	2.49 (.0064)	3.37 (.00037)	3.17 (.0008)	4.37 (.0000063)
Rounded-down data	2.42 (.0078)	3.45 (.00028)	3.10 (.0010)	4.44 (.0000046)

Study Size and Reliability of Contrast

Twelve out of the 46 studies (26%) contained fewer than 30 trials; 5 (11%) had fewer than 10. Small studies would be expected to give a less accurate estimate of any difference between direct hit and rank outcomes than relatively large studies. In case the presence of small studies had tended to obscure such a difference, in a post hoc analysis, each z was weighted by the number of trials in the study, and the overall z found separately for direct hits and for sums of ranks by the method described by Rosenthal (1991). The results are summarized in Table 2, for both rounded-up and rounded-down data. They indicate that the weighted Stouffer zs differ very little from the unweighted Stouffer zs given in

TABLE 2
WEIGHTED STOUFFER z VALUES FOR DIRECT HITS AND SUMS OF RANKS FOR GANZFELD STUDIES AND ALL STUDIES (WITH ONE-TAILED p VALUES GIVEN IN PARENTHESES)

	Ganzfeld studies (N = 42)		All studies (N = 46)	
	Hits	Ranks	Hits	Ranks
Rounded-up data				
Weighted Stouffer z	2.38 (.0087)	2.83 (.0023)	2.98 (.0014)	4.10 (.000021)
z_{diff}	0.32		0.79	
Rounded-down data				
Weighted Stouffer z	2.35 (.0094)	2.88 (.0020)	2.95 (.0016)	4.13 (.000018)
z_{diff}	0.37		0.83	

Table 1. The nonsignificant values of z_{diff} show that low sample size in some of the studies was not obscuring a relatively large difference between the outcomes of sums of ranks and direct hits.

Direct Hits Versus Ranks

Matched-pairs *t* tests were used to compare the standard normal deviates associated with each study's sum of ranks and number of direct hits and the effect sizes ($z/N^{1/2}$) associated with the two measures, both for ganzfeld studies alone and for the whole sample. The results are summarized in Table 3, which shows values of *t* obtained for the comparisons using rounded-up and rounded-down data. Although sums of ranks show superior performance in producing larger average *z*s and effect sizes, only one of the differences was statistically significant: the post hoc comparison for rounded-down data (which is conservative for hits) indicated that rank-based effect sizes were barely significantly larger (p = .049, two-tailed) than hit-based effect sizes for the whole database of studies, a result that would not withstand correction for multiple analysis.

TABLE 3

COMPARISON OF OUTCOMES DERIVED FROM DIRECT HITS AND SUMS OF RANKS FOR GANZFELD STUDIES AND ALL STUDIES

	Ganzfeld studies (N = 42)		All studies (N = 46)	
	Hits	Ranks	Hits	Ranks
Rounded-up data				
Mean z (*SD*)	0.38 (1.28)	0.52 (1.28)	0.47 (1.28)	0.64 (1.35)
	t_{diff} = 1.305, p = .20, 2-tailed		t_{diff} = 1.754, p = .09, 2-tailed	
Mean effect size (*SD*)	0.07 (0.26)	0.10 (0.28)	0.08 (0.25)	0.11 (0.28)
	t_{diff} = 1.683, p = .14, 2-tailed		t_{diff} = 1.781, p = .08, 2-tailed	
95% confidence limits	–0.01 to 0.15	0.02 to 0.18	0.01 to 0.15	0.03 to 0.19
Rounded-down data				
Mean z (*SD*)	0.37 (1.28)	0.53 (1.29)	0.46 (1.27)	0.65 (1.36)
	t_{diff} = 1.533, p = 0.13, 2-tailed		t_{diff} = 1.974, p = .054, 2-tailed	
Mean effect size (*SD*)	0.07 (0.26)	0.10 (0.28)	0.08 (0.25)	0.11 (0.28)
	t_{diff} = 1.734, p = .09, 2-tailed		t_{diff} = 2.019, p = .049, 2-tailed	
95% confidence limits	–0.01 to 0.15	0.02 to 0.18	0.01 to 0.15	0.03 to 0.19

Discussion

Although direct hits were overwhelmingly the main outcome measure of choice in this database, the results of the 46 studies examined indicate that sums of ranks are associated with greater statistical significance and higher effect sizes on average than direct hits but not to a statistically significant degree. It may be that there were too few studies to indicate a reliable difference, or it may be that there really is no difference. Examination of a larger database may clarify the situation, but, in the meantime, it would be premature to suggest that ranks should replace hits as the measure of choice because chance alone can easily account for the slight observed superiority of ranks.

The striking popularity of direct hits in the present database was due to the fact that the data consisted almost entirely of ganzfeld studies and all of the ganzfeld studies that reported either hits or ranks as their main outcome measure employed hits. The universality of hits as a measure for ganzfeld studies is likely to be due to their adoption by Honorton (Bem & Honorton, 1994) as the main measure for his influential autoganzfeld studies, which other researchers have sought to replicate. The results of the present meta-analysis suggest that in the absence of any demonstrable superiority of direct hits, researchers are more free than they might have supposed to use ranks instead, if they see advantages in doing so. For example, participants in free-response studies might find ranks more rewarding as a measure than direct hits, since they have less chance of absolute failure. Whether further empirical data will offer stronger arguments in favor of one measure or the other remains to be seen.

References

Bem, D. J., & Honorton, C. (1994). Does psi exist? Evidence for an anomalous process of information transfer. *Psychological Bulletin*, **115**, 4–18.

Bierman, D. J. (1995). The Amsterdam ganzfeld series III & IV: Target clip emotionality, effect sizes and openness. *Proceedings of Presented Papers: The Parapsychological Association 38th Annual Convention*, 27–37.

Bierman, D. J., Bosga, D. J., Gerding, H., & Wezelman, R. (1993). Anomalous information access in the Ganzfeld: Utrecht—Novice series I and II. *Proceedings of Presented Papers: The Parapsychological Association 36th Annual Convention*, 192–203.

Bosga, D. J., Gerding, J. L. F., & Wezelman, R. (1994). Target-affinity: An analysis of a psychological variable with possible implications for the ganzfeld procedure. *Proceedings of Presented Papers: The Parapsychological Association 37th Annual Convention*, 83–86.

BROUGHTON, R. S., & ALEXANDER, C. H. (1996). Autoganzfeld II: An attempted replication of the PRL ganzfeld research. *Proceedings of Presented Papers: The Parapsychological Association 39th Annual Convention*, 45–56.

CARPENTER, J. C. (1995a). An approach to assessing adjustment to the ganzfeld situation. *Proceedings of Presented Papers: The Parapsychological Association 38th Annual Convention*, 65–78.

CARPENTER, J. C. (1995b). Small group interaction as a medium for ESP response. *Proceedings of Presented Papers: The Parapsychological Association 38th Annual Convention*, 79–87.

DALTON, K. (1994). A report on informal ganzfeld trials and comparison of receiver/sender sex pairing: Avoiding the file drawer. *Proceedings of Presented Papers: The Parapsychological Association 37th Annual Convention*, 104–113.

DALTON, K., STEINKAMP, F., & SHERWOOD, S. J. (1996). A dream GESP experiment using dynamic targets and consensus vote. *Proceedings of Presented Papers: The Parapsychological Association 39th Annual Convention*, 57–72

HANSEN, G. P. (1986). A look at *z* scores (standardized ratings). In D. H. Weiner & D. L. Radin (Eds.), *Research in parapsychology 1985* (pp. 93–94). Metuchen, NJ: Scarecrow Press.

HANSEN, G. P., & UTTS, J. (1987). Use of both sum of ranks and direct hits in free-response psi experiments. *Journal of Parapsychology*, **51**, 321–335.

HONORTON, C. (1995). Impact of the sender in ganzfeld communication: Meta-analysis and power estimates. *Proceedings of Presented Papers: The Parapsychological Association 38th Annual Convention*, 132–140.

JOHANSSON, H., & PARKER, A. (1995). Replication of the ganzfeld findings: Using simplified ganzfeld procedure. *Proceedings of Presented Papers: The Parapsychological Association 38th Annual Convention*, 156–160.

KANTHAMANI, H., & BROUGHTON, R. S. (1992). An experiment in ganzfeld and dreams: A further confirmation. *Proceedings of Presented Papers: The Parapsychological Association 35th Annual Convention*, 59–73.

KANTHAMANI, H., & BROUGHTON, R. S. (1994). Institute for Parapsychology ganzfeld-ESP experiments: The manual series. *Proceedings of Presented Papers: The Parapsychological Association 37th Annual Convention*, 182–189.

KANTHAMANI, H., & KHILJI, A. (1990). An experiment in ganzfeld and dreams: A complimentary study. *Proceedings of Presented Papers: The Parapsychological Association 33rd Annual Convention*, 126–137.

KANTHAMANI, H., KHILJI, A., & RUSTOMJI-KERNS, R. (1988). An experiment in ganzfeld and dreams with a clairvoyance technique. *Proceedings of Presented Papers: The Parapsychological Association 31st Annual Convention*, 412–423.

KANTHAMANI, H., & PALMER, J. (1993). A ganzfeld experiment with "subliminal sending." *Journal of Parapsychology*, **57**, 241–257.

LANTZ, N. D., LUKE, L.W., & MAY, E.C. (1994). Target and sender dependencies in anomalous cognition experiments. *Journal of Parapsychology*, **58**, 285–303.

MCDONOUGH, B. E., DON, N. S., & WARREN, C. A. (1994). EEG in a ganzfeld psi task. *Proceedings of Presented Papers: The Parapsychological Association 37th Annual Convention*, 273–283.

MILTON, J., & STEVENS, P. (1997). An extended table and computer program for exact sum of ranks probabilities. *Proceedings of Presented Papers: The Parapsychological Association 40th Annual Convention*, 240–262.

MILTON, J., & WISEMAN, R. (1997). Ganzfeld at the crossroads: A meta-analysis of the new generation of studies. *Proceedings of Presented Papers: The Parapsychological Association 40th Annual Convention*, 267–282.

MORRIS, R. L., CUNNINGHAM, S., MCALPINE, S., & TAYLOR, R. (1993). Toward replication and extension of autoganzfeld results. *Proceedings of Presented Papers: The Parapsychological Association 36th Annual Convention*, 177–191.

MORRIS, R. L., DALTON, K., DELANOY, D. L., & WATT, C. (1995). Comparison of the sender/no sender condition in the ganzfeld. *Proceedings of Presented Papers: The Parapsychological Association 38th Annual Convention*, 244–259.

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

ROSENTHAL, R. (1991). *Meta-analytic procedures for social research.* London: Sage.

SCHOUTEN, S. A., & CAMFFERMAN. (1988). Evaluating free-response rating data. In D. H. Weiner & R. L. Morris (Eds.), *Research in parapsychology 1987* (pp. 117–121). Metuchen, NJ: Scarecrow Press.

SOLFVIN, G. F., KELLY, E. F., & BURDICK, D. S. (1978). Some new methods of analysis for preferential-ranking data. *Journal of the American Society for Psychical Research*, **72**, 93–110.

SPOTTISWOODE, S. J. P. (1993). Effect of ambient magnetic field fluctuations on performance in a free response anomalous cognition task: A pilot study. *Proceedings of Presented Papers: The Parapsychological Association 36th Annual Convention*, 143–156.

STANFORD, R. G. (1990, January). Secretary's report, 1989. PA News & Annual Report. Durham, NC: The Parapsychological Association, Inc.

STANFORD, R. G., & MAYER, B. (1974). Relaxation as a psi-conducive state: A replication and exploration of parameters. *Journal of the American Society for Psychical Research*, **68**, 182–191.

UTTS, J. (1986). The ganzfeld debate: A statistician's perspective. *Journal of Parapsychology*, **50**, 393–402.

WILLIAMS, C., ROE, C. A., UPCHURCH, I., & LAWRENCE, T. R. (1994). Senders and geomagnetism in the ganzfeld. *Proceedings of Presented Papers: The Parapsychological Association 37th Annual Convention*, 429–438.

WILLIN, M. J. (1996a). A ganzfeld experiment using musical targets. *Journal of the Society for Psychical Research*, **61**, 1–17.

WILLIN, M. J. (1996b). A ganzfeld experiment using musical targets with previous high scorers from the general population. *Journal of the Society for Psychical Research*, **61**, 103–108.

WISEMAN, R., HARRIS, P., & MIDDLETON, W. (1994). Luckiness and psi: An initial study. *Journal of the Society for Psychical Research*, **60**, 1–15.

ZINGRONE, N. L., HANSEN, G. P., & PERLSTROM, J. (1985). Series One and Two of a ganzfeld pilot study [Abstract]. *Journal of Parapsychology*, **49**, 263.

Department of Psychology
University of Edinburgh
7 George Square
Edinburgh EH8 9JZ
Scotland, UK

Appendix

Direct Hit and Sum of Rank Outcomes for Individual Studies

Study	Sole main analysis (hits or ranks)	N trials	Set size	Direct hits			Sum of ranks		
				Score	z	$z/N^{1/2}$	Score	z	$z/N^{1/2}$
GANZFELD ($N = 44$)									
Bem & Honorton '94 Pilot Series 1	Hits	22	4	8	0.99	0.21	46	1.62	0.35
Bem & Honorton '94 Pilot Series 2	Hits	9	4	3	0.25	0.08	23	−0.29	−0.10
Bem & Honorton '94 Pilot Series 3	Hits	35	4	10	0.32	0.05	78	1.36	0.23
Bem & Honorton '94 Series 101	Hits	50	4	12	−0.30	−0.04	128	−0.44	−0.06
Bem & Honorton '94 Series 102	Hits	50	4	18	1.60	0.23	117	0.95	0.13
Bem & Honorton '94 Series 103	Hits	50	4	15	0.67	0.09	117	0.95	0.13
Bem & Honorton '94 Series 104	Hits	50	4	18	1.60	0.23	105	2.48	0.35
Bem & Honorton '94 Series 105	Hits	6	4	4	1.78	0.73	8	2.47	1.08
Bem & Honorton '94 Series 201	Hits	7	4	3	0.69	0.26	16	0.33	0.12
Bem & Honorton '94 Series 301	Hits	50	4	15	0.67	0.09	119	0.69	0.10
Bem & Honorton '94 Series 302	Hits	25	4	16	3.93	0.79	42	3.72	0.74
Bierman '95 Amsterdam Series III	Hits	38[a]	4	16	2.15	0.35	79[b]	2.26	0.37
Bierman '95 Amsterdam Series IV	Hits	36	4	13	1.33	0.22	78[b]	1.72	0.29
Bierman et al. '93 Utrecht Novice Series I	Hits	50	4	13	0.03	0.00	120	0.57	0.08
Bierman et al. '93 Utrecht Novice Series II	Hits	50	4	12	−0.30	−0.04	131	−0.82	−0.12
Bosga et al. '94 Utrecht Experienced Series I	Not reported	50	4	10	−0.98	−0.14	121	0.44	0.06
Broughton & Alexander '96 FT1	Hits	50	4	12	−0.30	−0.04	135	−1.33	−0.19

Study	Sole main analysis (hits or ranks)	N trials	Set size	Direct hits: Score	Direct hits: z	Direct hits: $z/N^{1/2}$	Sum of ranks: Score	Sum of ranks: z	Sum of ranks: $z/N^{1/2}$
Broughton & Alexander '96 FT2	Hits	50	4	9	−1.33	−0.19	130	−0.69	−0.10
Broughton & Alexander '96 EC1	Hits	51	4	19	1.81	0.25	118	1.13	0.16
Broughton & Alexander '96 CLAIR1	Hits	50	4	11	−0.64	−0.09	131	−0.82	−0.12
Broughton & Alexander '96 GEN1	Hits	8	4	3	0.46	0.16	18	0.47	0.17
Carpenter '95a/Zingrone et al. '85 Study 1	Not reported	30	4	9	0.45	0.08	75	−0.08	−0.01
Carpenter '95a/Zingrone et al. '85 Study 2	Not reported	20	4	4	−0.75	−0.17	52	−0.50	−0.11
Dalton '94	Hits	29	4	12	1.76	0.33	59	2.17	.40
Johansson & Parker '95 Study 1[c]	Hits	30	4	6	−0.83	−0.15	69	0.90	0.16
Johansson & Parker '95 Study 2[c]	Hits	30	4	11	1.25	0.23	67	1.22	0.22
Johansson & Parker '95 Study 3[c]	Hits	30	4	11	1.25	0.23	63	1.88	0.34
Kanthamani & Broughton '94 Series 1	Neither	30	4	6	−0.83	−0.15	87	−2.05	−0.37
Kanthamani & Broughton '94 Series 2	Not reported	40	4	12	0.57	0.09	96	0.49	0.08
Kanthamani & Broughton '94 Series 3	Not reported	40	4	8	−0.91	−0.14	97	0.35	0.06
Kanthamani & Broughton '94 Series 4	Not reported	65	4	24	2.01	0.25	151	1.22	0.15
Kanthamani & Broughton '94 Series 5a	Neither	4	4	1[d]	−0.48	−0.24	9	0.22	0.11
Kanthamani & Broughton '94 Series 5b	Neither	10	4	1[d]	−1.59	−0.50	31	−1.85	−0.59
Kanthamani & Broughton '94 Series 6	Neither	66	4	18	0.31	0.04	162	0.27	0.03
Kanthamani & Broughton '94 Series 7[c]	Not reported	–	–	–	–	–	–	–	–
Kanthamani & Broughton '94 Series 8[c]	Not reported	–	–	–	–	–	–	–	–
Kanthamani & Palmer '93	Hits	22	4	2[d]	−2.17	−0.46	63[d]	−1.62	−0.35
McDonough et al. '94	Hits	20	4	8[b,d]	1.27	0.28	45[b,d]	0.90	0.20
Morris et al. '93 Cunningham Study	Hits	32	4	13	1.78	0.31	71	1.34	0.24

Study	Sole main analysis (hits or ranks)	*N* trials	Set size	Direct hits Score	Direct hits z	Direct hits $z/N^{1/2}$	Sum of ranks Score	Sum of ranks z	Sum of ranks $z/N^{1/2}$
Morris et al. '93 McAlpine Study	Hits	32	4	8	–0.17	–0.03	80	–0.08	–0.01
Morris et al. '95	Hits	97	4	32	1.67	0.17	227	1.36[f]	0.14
Williams et al. '94	Hits	42	4	5	–2.30	–0.35	119	–2.01	–0.31
Willin '96a	Hits	100	4	24	–0.33	–0.03	254	–0.31[f]	–0.03
Willin '96b	Hits	16	4	4	–0.24	–0.06	40	–0.11	–0.03
NON-GANZFELD ($N = 10$)									
Carpenter '95b[g]	Neither	286	4	81	1.22	0.07	672	2.25[f]	0.13
Dalton et al. '96	Hits	32	4	13	2.51	0.44	57	3.72	0.66
Kanthamani & Broughton '92 Dream Trials[c]	Neither	–	–	–	–	–	–	–	–
Kanthamani & Khilji '90 Dream Trials[c]	Neither	–	–	–	–	–	–	–	–
Kanthamani et al. '88 Prelim. Dream Trials[c]	Neither	–	–	–	–	–	–	–	–
Kanthamani et al. '88 Pilot Dream Trials[c]	Neither	–	–	–	–	–	–	–	–
Lantz et al. '94 1992 Study[h]	Ranks	–	–	–	–	–	–	–	–
Lantz et al. '94 1993 Study[h]	Ranks	–	–	–	–	–	–	–	–
Spottiswoode '93[i]	Ranks	101	5	22	0.35	0.03	303	0.00[f]	0.00
Wiseman et al. '94	Ranks	114	4	35	1.29	0.12	263	1.80[f]	0.17

[a]Computer records of rating data were only kept for 38 of the 40 trials reported by Bierman (1995), so the comparison of direct hits and sum of ranks here is restricted to those 38 trials.

[b]An average score of an integer plus a half has been rounded up to the nearest integer.

[c]Although not apparent from the interim-report paper, this experiment consisted of three separate studies whose data are presented here (personal communication, Parker, 1997).

[d]Scores are based on averages of two or more judges.

[e]Data from these studies were not made available for the present paper because Kanthamani intends to publish her own comparisons of outcome measures separately.

[f]Z was calculated using the normal approximation to the sum of ranks distribution.

[g]Since publication of the interim report on this study, Carpenter has completed further trials which are reported here.

[h]Data from these studies were not made available.

[i]Since publication of the interim report on this study, Spottiswoode has completed further trials which are reported here.