DOCUMENT RESUME

ED 411 294 TM 027 307

AUTHOR Gray, William M.; Fox, Christine M.
TITLE Rasch Scaling of Different Forms of Thought and Different Logical Operations.
PUB DATE 1997-03-00
NOTE 54p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Child Development; *College Students; Developmental Stages; Higher Education; Item Response Theory; *Logical Thinking; *Piagetian Theory; *Scaling; Test Items; *Thinking Skills
IDENTIFIERS Partial Credit Model; *Rasch Model

ABSTRACT

Twenty-four Piagetian-based written items representing different forms (stages) of thought and different logical operations were solved by 553 participants from required introductory logic classes in a state university in a medium-sized midwestern city and from a suburban school district of that city. Responses to each item were scored by quality of thought demonstrated by the participants according to guidelines provided by B. Inhelder and J. Piaget in various sources. A partial credit Rasch analysis produced an item ordering that was theoretically supportable. Comparison of the item ordering from this study with the item ordering from a different study supported the view that Piagetian-based tasks can be scaled with sample independence and Rasch analysis can produce sample independent scales. (Contains 3 tables, 6 figures, and 76 references.) (Author/SLD)

# Rasch Scaling of Different Forms of Thought and Different Logical Operations



William M. Gray and Christine M. Fox

University of Toledo



Paper presented at the annual meeting of the American Educational Research Association, Chicago, March 25, 1997.

We thank Biological Sciences Curriculum Study (BSCS) for allowing unlimited use of How Is Your Logic? which was created by the first author for BSCS with partial support from the National Science Foundation. Preparation of this paper was supported, in part, by a sabbatical leave granted to the first author by The University of Toledo. The existence of this paper was initiated because of the quiet, patient, insistence of Trevor Bond who convinced the first author of the quality of the data and the possibilities inherent in Rasch scaling analysis. We also thank Richard J. Hofmann for his clear insight in helping us resolve some issues in our analysis.

Correspondence concerning this paper should be addressed to William M. Gray, Educational Psychology, University of Toledo, Toledo, Ohio 43606-3390. Electronic mail may be sent via Internet to William.Gray@utoledo.edu.

# Rasch Scaling of Different Forms of Thought and Different Logical Operations

## Abstract

Twenty-four, Piagetian-based written items representing different forms (stages) of thought and different logical operations were solved by 553 participants from required introductory logic classes in a state university in a medium-size midwestern city and from a suburban school district of that city. Responses to each item were scored by quality of thought demonstrated by the participants according to guidelines provided by Inhelder and Piaget in various sources. A partial credit Rasch analysis produced an item ordering that was theoretically supportable. Comparison of the item ordering from this study with the item ordering from a different study supported the view that Piagetian-based tasks can be scaled with sample independence and Rasch analysis can produce sample independent scales.

Rasch Scaling of Different Forms of Thought and Different Logical Operations

An assumption of Piaget's biological approach to psychological development is that development is continuous. Through continuous adaptations a hierarchy of patterns of functioning become relatively stable and this hierarchy is represented by a sequence of forms (stages) of thought (i.e., sensorimotor, preoperations, concrete operations, formal operations).

Traditionally, this sequence of forms has been assessed by tasks representative of one of the forms, and then evaluating individuals' responses to the tasks, especially responses that are not adequate for successful solving of the tasks. For example, except for conservation of occupied volume, the classic conservation tasks require concrete operations for success and are used to distinguish between preoperations and concrete operations (Piaget & Inhelder, 1941/1974). Inhelder's classic, simple, inductive physics tasks (Inhelder & Piaget, 1955/1958) require formal operations for success and are used to distinguish among preoperations, concrete operations, and formal operations. Since the rediscovery of Piaget's ideas by American psychologists in the late 50s and early 60s (Kessen, 1996), most replications, extensions, etc., of Piaget's ideas (e.g., see Blasi & Hoeffel, 1974; Farrell, 1969; Lawson, 1985; Neimark, 1975, 1979, 1982, for reviews on formal operations) have focused on the demographic characteristics of participants, whether the various adaptations can be taught, whether the various adaptations can be accelerated, does the nature of the tasks' directions or the content of the tasks themselves affect success, etc. However, focusing on the participants and their demographic, psychological, or sociological characteristics, or the impact of various environmental manipulations, was not the intent or focus of Piaget's perspective. Rather, "over the series of [Piaget's] works which attack intelligence at different points on the growth continuum [i.e., forms of thought] and focus on different functions [emphasis added by us], the over-all aim has been to trace the development of intelligence as it comes to deal with increasingly complex problems or as it deals with simple problems in increasingly more efficient ways" (Parsons, 1958, p. xi). "From the beginning, Piaget's stages were meant as a classification of forms of thinking, not of individual children" (Chapman, 1988, p. 33). And, in Piaget's own words, "the task still remains--and we are occupied with it at the moment--of studying the formation of reason in the child in and for itself" (Piaget, 1931/1955, p. 24) and

> the object of these studies, initially, was not to establish a scale of development and to obtain precise determinations of age as regards stages. It was a question of trying to understand the intellectual mechanism used in the solution of problems and of determining the mechanism of reasoning. . . . I do not despair of obtaining a scale which will perhaps not be exact within three months but whicvh, from the very fact that it will give a structured whole, will reveal more than would a system of compensations such as the mosaic-tests. (Piaget, 1956, pp. 89–90)

Although these statements were written in the early and mid part of Piaget's research efforts, their message continued to be one of the major focal points, if not the major focal point, of Piaget's oeuvre. In essence, Piaget's focus and emphasis was on investigating different forms of thought and their hierarchical complexity (Claparède, 1931/1955), not on classifying children. And, if the forms of thought formed a scale, that was also acceptable.

Recently, there has been a renewed interest in the hierarchical complexity of Piaget's forms of thought (Bond, 1989, 1995a, 1995b; Bond & Bunting, 1995; Demetriou, Efklides, Papadaki, Papantoniou, & Economou, 1993; Demetriou, Efklides, Platsidou, 1993; Gray, Bond, Fox, & Hofmann, 1996; Gray & Fox, 1996; Hautamäki, 1989; Mueller, Reene, & Overton, 1994; Mueller, Winn, Overton, 1995; Noelting, Coudé, & Rousseau, 1995; Noelting, Rousseau, & Coudé, 1994; also, see a dissertation by Docherty, 1973, which preceded this renewal by 20 years). These attempts all use a scaling approach created by Rasch (1980; Wright & Masters, 1982; Wright & Stone, 1979) which tests for an underlying dimension that conforms to an item characteristic curve represented as an ogive. Part of the power of the approach is that the Rasch model performs a log transformation on the raw data which results in a transformation from ordinal level data to interval level data. This transformation is unique in that it results in specific objectivity which refers to a test in which the measurement of a person's trait is independent of the set of items used to measure that trait and item calibration is independent of the set of people who take the test.

Specific objectivity allows the Rasch model to overcome problems with classical psychometric theory. With classical theory, item difficulties, discriminations, and ability levels are confounded because item difficulty is defined as the percentage of correct responses to an item for a given sample. Likewise, person ability is defined as the percent of items answered correctly. Thus, decisions about a person's ability or an item's difficulty are confounded with the specific set of items taken and the specific sample measured. Rasch calibrations and measures, on the other hand, are absolute across various samples (sample free measurement) and test items (item free calibration) within standard error estimates. In other words, independent random samples of items that have been calibrated using the Rasch model will yield equivalent ability estimates for any person or group from the population (Snyder & Sheehan, 1992). Therefore, the Rasch model contains no bias because it unconfounds measurement and calibration, and hence an estimate of a person's ability on a given construct is not influenced by the difficulty of the particular test item taken.

Using a textword search for Rasch on OhioLINK's Ovid version of PsyINFO (1967–present), we found 448 sources, with only three sources (Bond, 1995a, 1995b; Docherty 1973) focusing on a Piagetian approach to thought. None of these three sources tested sample independence or item independence. Similar searches of three other databases on OhioLINK produced the following: (a) on H. W. Wilson's version of Education Index (1983–present), we found 54 sources, but none focusing on a Piagetian approach to thought. (b) On Dissertation Abstracts (1986–present), we found 267 sources, with only Wilson (1984) and the aforementioned Docherty (1973) dissertation using Piagetian data. Docherty focused on scaling Piagetian-type tasks based on data generated from children. Wilson (1984) used a Monte Carlo approach to test his Saltus psychometric model of hierarchical development. And (c) on Ovid's version of ERIC (1966–present), we found 675 sources, with only two citing Piaget (Cornish & Wines, 1980; Elliott, 1982), but neither really focusing on Piaget's ideas. A cursory review of the titles and abstracts of the 1444 sources suggested that none tested the sample independence and/or item independence aspects of Rasch scaling analysis, especially with theoretically-based, empirically-generated data. If such research was reported, it was not obvious nor did it utilize data from a Piagetian perspective. Having said this, we must issue some caveats: First, Wilson (personal communication, March 13, 1997) provided more detailed information than that provided in an electronic abstract (Wilson, 1984) including the facts that his dissertation used empirically generated data and investigated item independence and sample independence. He found that there was evidence for item and sample dependence in fairly consistent patterns. Second, three works cited by Bond (1995a) (i.e., Andrich & Constable; Hacker, Pratt, & Matthews; Hautamäki, 1989) were not included in any of the databases. The Hautamäki (1989) omission is an especially egregious error as it is a report of Finnish early adolescents who completed a Finnish translation of Shayer's, English language, group-administered, science reasoning tasks, which were based on Piaget and Inhelder's original, French Swiss, concrete, physical tasks. Hautamäki reports a remarkable consistency with Shayer, Küchemann, and Wylam's (1976) results and provides powerful support for the Rasch contention of sample independence. The words of the discussion reporter are clear and powerful (Matthews, 1989, p. 350):

> In an animated discussion which followed, it was thought remarkable that a theory originating in Geneva, modified by Shayer in Britain and then translated into the Finnish enironment [sic] should still show unidimensionality, as shown in the

striking agreement shown in figure 14.3.1 [Hautamäki, 1989, p. 345]. In addition, it was noted that Rasch modelling [sic] showed equal differences between stages 2B [Concrete II] and 3A [Formal I], and between stages 3A [Formal I] and 3B [Formal II]. Such robustness and consistency of test with theory is unusual among psychological models, it was argued: so why do so many people consider it invalid?

To some, the absence of discontinuities is a scandal; it disturbs them that there are persons with intermediate scores on tests like 2B/3A, implying continuous development. [For a very interesting statement on this phenomenon see Inhelder, 1956, pp. 78–84 and note that it was made during a conference in 1953.] It was also argued that the very success of Piaget's work has led to many simplistic versions of his theories, and many group tests being too far removed from the original Genevan research.

Thus, although it is possible that sample independence and/or item independence have been tested, except for Hautamäki (1989) and Wilson (personal communication, March 13, 1997), we know of no other study that has used theoretically-based data generated from real participants to test the sample independence and/or item independence aspect of Rasch scaling analysis.

In a previous study (Gray & Fox, 1996), we conducted an ex post facto partial credit Rasch analysis of the written responses to a set of 24 Piagetian-based, written items which were collected on 577 junior high school students who participated in an innovative science curriculum. Mean age was 14.08 (SD = .52) with five participants not providing complete birthdate information; males were 47.31% of the sample. Those data used a theoretically constructed stage scoring model of 12345678 (1 = "too difficult," "can't do," did not respond, etc.; 2 = Preoperations; 3 = Preoperations–Concrete I; 4 = Concrete I; 5 = Concrete II; 6 = Concrete II–Formal I; 7 = Formal I; & 8 = Formal II) to evaluate participants' answers to the items. Person separation was 2.41 and item separation was 15.49. This theoretically-based stage scoring model provided for a refined differentiation between concrete operations and formal operations as well as for a refined differentiation between subforms (i.e., substages) within each form of thought. For this homogeneous sample, empirical placement of the items on the unidimensional scale was excellent for the differentiation between concrete operations and formal operations, excellent for the differentiation between the concrete operations of multiplication of classes and addition of asymmetrical relations, and, depending on the operation, adequate to excellent for the differentiation among the various formal operations. The greater complexity within formal operations than within concrete operations is not unexpected as the participants were of an age (early to mid adolescence) when concrete operations should be clearly differentiated, consolidated, stabilized, and integrated, whereas formal operations are beginning to be manifested and, thus, their differentiation, consolidation, stabilization, and integration should not be as clear (Gray, 1990; Inhelder & Piaget, 1955/1958, 1959/1969; Martorano, 1975; Nassefat as described in Flavell & Wohlwill, 1969, pp. 96–98; Piaget, 1947/1966, pp. 139–153, 1953/1957).

The present study was an ex post facto replication of Gray and Fox (1996) with a more age heterogeneous sample. Thus, we tested the sample independence aspect of Rasch scaling analysis with theoretically-based data generated from real participants.

## Method

### Participants

Participants (N = 553) were from a suburb of a medium-size, midwestern city and required introductory logic classes in the state university within the city. Mean age was 16.66 (SD = 4.97), ranging from 10.95 to 48.46. Males were 55.3% of the sample. They were part or a larger study of 746 participants that focused on the relations between moral reasoning and operational thought. For the complete sample, mean age was 16.47 (SD = 5.20) ranging from 9.48 to 48.46 with 55.5% being males. The reduced sample size reported here occurred because scheduling problems prevented 193 participants from completing both forms of the test measuring operational thought.

Procedure

Within a two week period, participants solved 26 written items that represent different formal operations and different concrete operations. On the first day participants completed Form A of How Is Your Logic? (Gray, 1976a) and on a second day they completed Form B of How Is Your Logic? (Gray, 1976b). There was no time limit to complete each form of the test.

Instrument

Both Form A and Form B of How Is Your Logic? (HIYL) are composed of 13 items, eight measuring formal operations and five measuring concrete operations. Each item requires a constructed response. Answers are classified according to the form of thought required for a correct answer and the form of thought produced by participants on the items.

On each form, concrete operational items represent four of the eight concrete operational groupings: Grouping III--Bi-univocal Multiplication of Classes (Inhelder & Piaget, 1959/1969, pp. 151–195) by pairing all members of one set with all members of a second set, Grouping IV--Co-univocal Multiplication of Classes (Flavell, 1963, pp. 179–180; Piaget, 1947/1966, p. 45) by pairing all members of one set with the one member of a second set, Grouping V--Addition of Asymmetrical Relations (Inhelder & Piaget, 1959/1969, pp. 247–268; Piaget, 1941/1965, pp. 122–157) by seriating eight or nine figures by height, and Grouping VII--Bi-univocal Multiplication of Relations (Inhelder & Piaget, 19159/1969, pp. 269–279; Piaget, 1941/1965, pp. 96–121) by jointly seriating a set of 16 objects along two dimensions. Correct answers receive a score of Concrete I (i.e., 4) and incorrect answers receive a score of Preoperations-Concrete I (i.e., 3), Preoperations (i.e., 2), or Did Not Attempt, "too hard,", etc. (i.e., 1) depending on the quality of the incorrect answer. Scoring criteria are the characteristics of the various forms (e.g., Preoperations) and sub-forms of thought (i.e., Preoperations–Concrete I or Concrete I) specified in the aforementioned pages describing the concrete operations included on the test.

Formal operations items represent beginning formal operations (i.e., Formal I) and/or consolidated formal operations (i.e., Formal II) depending on the quality of thought necessary for solving the items. Operations include (a) make correct inclusions (Inhelder & Piaget, 1955/1958, pp. 67–79) by determining which one of three variables systematically varied across four conditions "determines" (i.e., co-occurs with) a fourth variable, (b) deny incorrect inclusions/make correct exclusions (Inhelder & Piaget, 1955/1958, pp. 67–79) by determining that neither one of two variables systematically varied across four conditions "determines" (i.e., does not co-occur with) a third variable, (c) combinatorial thought by generating all possible 15 non-null combinations of four items (Inhelder & Piaget, 1955/1958, pp. 107–122; Piaget & Inhelder, 1951/1975, pp. 161–172), (d) combinatorial thought by generating all possible 24 permutations of four items (Piaget & Inhelder, 1951/1975, 173–194), (e) proportional reasoning (Inhelder & Piaget, 1955/1958, pp. 164–181) by determining from which one of two bags of candy a specific type of two candies is most likely to be randomly drawn, and (f) proportional reasoning (Inhelder & Piaget, 1955/1958, pp. 164–181) by determining from which one of two garages, each containing cars which are one of two colors, a specific color car is most likely to randomly appear. Each make correct inclusion, deny incorrect inclusions/make correct exclusions, and proportional reasoning item requires a judgment as to the correct answer and a separate explanation of the judgment. This separate scoring for judgments and explanations provides for the opportunity to look at the effects of different criteria for determining form of thought, a topic that has been rather controversial through the years (e.g., see Bingham-Newman & Hooper, 1975; Brainerd, 1973, 1974, 1975a, 1975b, 1977, 1978; Chandler & Chapman, 1991; Pinard & Laurendeau, 1969; Reese & Schack, 1974). Correct answers to the make correct inclusion judgment and explanation items, the combination items, and the proportional reasoning candy item receive a score of Formal I (i.e., 7). Correct answers to the deny incorrect inclusions/make correct exclusions judgment and explanation items, the permutation items, and the proportional reasoning car item receive a score of Formal II (i.e., 8). Placement of the proportional reasoning items at Formal I or Formal II was based on work by Noelting (1976, personal communication, September 8, 1975). Correct answers to the proportional reasoning explanation items receive a score of Formal II (i.e., 8) or Formal I (i.e., 7) depending on the quality of explanation. For all questions, incorrect answers receive a score of Concrete II-Formal I (i.e., 6), Concrete II (i.e., 5), Concrete I (i.e., 4), Preoperations-Concrete I (i.e., 3), Preoperations (i.e., 2), or Did Not Attempt (1) depending on the quality of the incorrect answer. Scoring criteria are the characteristics of the various forms (e.g., Preoperations) and sub-forms of thought (e.g., Formal II) specified in the aforementioned pages describing the formal operations included on the test. Table 1 describes the distribution of operations and their representative items for both forms.

All items, concrete and formal, are scored conservatively. For example, on a make correct inclusion item (Formal I) if an answer clearly indicates Concrete II thinking and only partially indicates Formal I thinking, the answer is scored as Concrete II–Formal I. This conservative approach is congruent with the descriptions presented in various works by Piaget and Inhelder, and it also reduces false positives by requiring that an answer clearly represent a form or subform of thought for the answer to be classified as such.

Raters were doctoral students in educational psychology who were trained by the first author, who created the tests and scoring criteria. Each scorer was required to first score 25–30 tests and their scores for each item were then matched with the first author's scores for the items. Percent agreement between each rater's training item scores and the first author's scores was 85% to 100% with only six items having less than 100% agreement. (All of these items were the explanation items.) Scoring of each item was discussed, regardless of whether a scorer's rating matched the rating of the first author. Any discrepancies in scoring were resolved and appropriate clarification of the scoring criteria were made. Subsequently, the tests that were used as training tests were returned to the pool of tests to be scored and they were rescored along with the tests not used in training.

Analysis

A Rasch partial credit analysis was used. Although it is often assumed that items are written in an understandable way and that the number, type, and labeling of categories are relevant and meaningful (See Clark & Schober, 1992, for a discussion of how changes in item wording can produce unpredictable shifts in responses; also see the first footnote to Table 1 which describes a similar phenomenon where it was determined that the directions for two HIYL items are ambiguous.), as the analysis progressed, it became obvious that the original, theoretically-based, eight category stage scoring model may not produce the best results. We then proceeded to alter the category structure of the original stage scoring model by systematically collapsing the categories. This process continued using four criteria to determine the best stage scoring model: (a) high item and person separation; (b) step difficulties should increase (i.e., monotonically increasing) and have "steep gradients" within an item; and (c) category usage should not be idiosyncratic or any idiosyncrasy should be relatively stable (i.e., fit statistics should be consistent or improved). These three criteria are statistical and stem from Lopez's (1996) work. The fourth criterion is theoretical (Inhelder, 1956; Piaget, 1960; Pinard & Laurendeau, 1969) and suggests that the ordering of the items be considered when deciding between statistically comparable stage scoring models. This means that if two or more stage scoring models produce similarly acceptable results, the ordering of the items within each stage scoring model be considered in determining the best solution. We, then, analyzed 17 theoretically plausible stage scoring model recategorizations using BIGSTEPS (Linacre & Wright, 1991–1994).

# Results

## Reliability

Coefficient α was .82 for the eight concrete items and 16 formal items together, .83 for the formal items, and .57 for the concrete items. The low α for the concrete items was a function of a ceiling effect, which was not unanticipated given that adults composed part of the sample.

Criterion one was used to determine reasonableness of item separation and person separation of the 17 stage scoring models. Figure 1 represents the joint item and person separation values of the nine stage scoring models which yielded the highest person and item separations. Of those nine stage scoring models, one six category model (11233456) and one seven category model (11234567) were the best. The original eight category model (12345678) had the highest person separation (2.13), but the second lowest (13.14) item separation. The two best models were then compared with respect to fit standard deviation, that is, how much variability each displayed in terms of idiosyncratic responses.

Figure 2 for item fit shows that, compared to the other seven models, the infit and outfit values of the six category model (11233456) and the seven category model (11234567) were not exceptional. Both models had identical infit and outfit values with the outfit value (0.29) being moderate and the infit value (0.12) being tied for the largest value. Figure 3 for person fit shows that the infit value was slightly better for the seven category model (11234567) than the six category (11233456) model and both had the same outfit value. Compared to the other stage scoring models, these two models had reasonably low infit and outfit values.

Criterion two was assessed by looking at differences between category measures to determine if scores were increasing from category to category and also to determine the size of the difference in scores between adjacent categories. Table 2 lists the average category measures and category fit measures for the two best scoring models. In terms of monotonicity, the six category model (11233456) only had 3 instances (3/62 = 4.84%) where there was no monotonicity in moving from a lower category to the next higher category (B4 moving from 3 to 4, B5 moving from 2 to 3, & B11 moving from 1 to 3). The seven category model (11234567) had 13 instances (13/76 = 17.11%) where there was no monotonicity in moving from a lower category to the next higher category. The six category model also had very good gradients in moving from a lower category to the next higher category. Clearly, the six category model produced better results.

Criterion three was assessed by looking at idiosyncratic use of categories. Table 2 also presents measures of idiosyncratic use of categories as evidenced by the infit MS and outfit MS for both stage scoring models. Mean square fit statistics have an expected value of 1. Values greater than 1 indicate "noise" or an idiosyncratic use of the category. Although both scoring models yielded noisy categories (≥ 1.2), the fit was much better for the six category scoring model than for the seven category model as the six category scoring model had 18.6% (16/86) infit values and 17.44% (15/86) outfit values ≥ 1.2, whereas the seven category scoring model had 21% (21/100) infit values and 24% (24/100) outfit values ≥ 1.2. Clearly, the six category model had less idiosyncratic category errors.

Both stage scoring models were comparable in terms of item and person separation. The six category model had better monotonicity between categories within items and its step gradients were generally reasonable. Also, the six category model was clearly better in lack of idiosyncratic use of its categories. Finally, the item ordering was identical for both models. Thus, the six category model was selected as the best representation of the data.

Validity

Order validity was examined through the difficulty ordering of the items and the diagnosis of misfit. Item measures and item fit for the six category model are presented in Table 3 and a map of persons and items on the same logit scale is presented in Figure 4.

Person reliability was ..81 and person separation was 2.08, which suggests that the items separated the participants into two statistically distinct groups, although where the separation into two groups of persons should occur is not clear in Figure 4. Item reliability was 1.00 and item separation was 14.87, which suggests that this group of persons can separate the items into 14 distinct groups, although Figure 4 suggests 6 or 7 groups of items, not 16. In addition, in Table 3, BIGSTEPS arbitrarily rearranges the order of some items compared to their order based on measure reported in Table 3. For example, B! and B5 are listed in measure order sequence, but A! and A$ are reversed. A9 and Be have switched positions and B11, A12, and A6 have shifted positions by being placed in a an earlier position and the easiest item is placed last on the list. In addition, items A1 and A4 are separated by .05 logits and are printed on the same line whereas A@ and B9 are separated by .03 logits and are printed on separate lines.

In Table 3 the horizontal gaps differentiate between items representing different operations and the horizontal double rule separates the concrete operational items from the formal operational items. Item placements clearly separated the concrete items and the formal items with no concrete item being more difficult than a formal item and no formal item being easier than a concrete item.

Within the concrete items, the addition of asymmetrical relations items (B1, B5, A4, A1) were the easiest and the multiplication of classes items were more difficult. This was similar to our previous work (Fox & Gray, 1997; Gray & Fox, 1996) which found that addition of asymmetrical relations items were easier than multiplication of classes items.

Within the formal items, the combination items were the easiest and the proportional reasoning explanation items (A13, B12) were the hardest. This is consonant with similar results reported by Gray and Fox (1996), Fox and Gray (1997), and Martorano (1977) who found combinatorial operations easier than inclusion/exclusion and proportional reasoning, but contrary to that suggested by Kuhn and Angelev (1976) who found inclusion/exclusion (pendulum problem) easier then combinatorial thought (chemicals problem). In addition, the combinatorial items and the proportional reasoning explanation items were clearly separated from the other formal items. Based on theory (Inhelder & Piaget, 1955/1958: Piaget & Inhelder, 1951/1975) and empirical data (Martorano, 1977), it was also expected that the permutation items would be more difficult than the combination items; this was the case. However, the data regarding the inclusion items (A2, A3, B9, B10) and the exclusion items (A9, A10, B3, B4) were more complicated than theoretically (Inhelder & Piaget, 1955/1958) and empirically expected (Hubbs-Tait, 1986; Kuhn & Brannock, 1977; Kuhn & Ho, 1977; Kuhn, Langer, Kohlberg, & Haan, 1977; and Commons & Kuhn as cited in Kuhn & Phelps, 1979). As expected, the make correct inclusion judgment items (A2, B9) were the easiest. The remainder of the inclusion items (A3, B10) and exclusion items (A9, A10, B3, B4) were somewhat out of order. However, considering the error rate associated with each measure, the variation in order is not unreasonable. Finally, the expectation or the position of the proportional reasoning judgment items was that, theoretically (Inhelder & Piaget, 1955/1958) and empirically (Martorano, 1977), they should be more difficult than the permutation items because the logic underlying the proportional reasoning judgment items requires the INRC group, whereas the permutation items do not. This expectation was not supported. However, the proportional reasoning judgment items were distinguished as misfitting items. Their misfit may be a function of them being the only INRC-based items or it may be a function of them being the only two items requiring mathematical computation for a correct answer, although the mathematical computations required for success on the items requires the logic of the INRC group unless the numerical answers are produced without any understanding. This difference between correct judgment and lack of understanding (explanation) is supported by the distance between the proportional reasoning judgment items (A12, B11) and the proportional reasoning explanation items (A13, B12).

The preceding information leads to the focus of the study: the sample independence of the items through Rasch scaling. If sample independence occurs, then the value of the item measures should be comparable when two different samples are used to scale the items. That is, it would not be expected that item measures for the same item should be different when the measures are generated from two different samples. Figure 5 and Figure 6 directly address this issue. In looking for theoretical and empirical explanations for misfitting items, Fox and Gray (1997) conducted an analysis similar to this study with the same 24 items. Whereas our sample was heterogeneous with respect to age (range = 10.95–48.46, M = 16.66, SD = 4.97), Fox and Gray's sample was restricted to eighth graders (range = XXXXXXXX, M = 14.08, SD = .52). The same process and criteria were used in both studies to determine the optimal scaling of the items except that Fox and Gray did not consider the fourth criterion regarding theoretical ordering of the items when determining the optimal solution. Rather, they let the scaling process determine the optimal solution. They found two solutions that were almost identical in being the optimal solution. One solution was the same six category stage scoring model (11233456) found to be optimal in this study. A second solution was a four category stage scoring model (11233444). Given that the only difference between the two solutions is a differentiation among subforms of thought representative of formal operations (Concrete II–Formal I, Formal I, Formal II, for the six category scoring model vs. Formal, for the four category scoring model), it is not surprising that a correlation between the item measures for both solutions was 1.00. What is more interesting and striking is the correlation ($r = .984$) between the item measures from the six category scoring model from this study with the item measures from the six category scoring model from Fox and Gray (1997). In addition, the correlation ($r = .978$) between the item mesas from the six category scoring model from this study with the item mesas from the four category scoring model from Fox and Gray (1997) was also exceedingly high. It is important to reemphasize the fact that our sample consisted of late elementary school, junior high, high school, and college undergraduates from medium size midwestern urban university whereas Fox and Gray's sample was from eighth graders participating in an innovative science program throughout the nation. Figure 5 and Figure 6 provide the joint plot of the item measures for the two six category scoring models and for the six category model with the four category model, respectively. As expected, both plots form almost a straight line.

## Discussion

The results were excellent and provide for reflection on several issues: First, it appears that Piagetian-based, written items measuring cognitive development can be used to replicate some of the "standardization" of the Genevans that is variously cited (e.g., Inhelder, 1971; Piaget, 1956), but difficult to verify and/or replicate. Although using different items, different samples, and different scoring techniques (Shayer, Küchemann, & Wylam, 1976, and Hautamäki, 1989, used a 1/0 scoring scheme whereas we used a differentiated scoring scheme intended to represent the various subforms of incorrect reasoning), our Figures 5 and 6 are as striking as Hautamäki's (1989) almost perfect duplication of Shayer, Küchemann, and Wylam's (1976) results. These two sets of results clearly indicate that well designed Piagetian-based tasks can be standardized, thus, eliminating some of the continuing criticism that the Genevan results can not be replicated.

This leads to a second issue which is that good theory leads to good instruments/items/etc., which lead to good data that needs a good analytic technique. Assuming that the items of How Is Your Logic? are faithful representations of the logical problems used by Inhelder and Piaget to assess cognitive development, and that similar results can be generated in the future, then it is clear that Piagetian theory provides a rich theoretical framework from which to understand the results. Bond (1995, 1995b; Bond & Bunting, 1995), Shayer, Adey, and Wylam (1981), as well as many others, have stressed the idea that Piagetian theory provides an enormously rich, but sophisticated and complicated, framework from which to investigate cognition. Although this is generally acknowledged, the lack of accuracy, precision, and standardization involved in many attempts at using a Piagetian framework have placed it in a less favorable light than when it became popular during the 60s and 70s.

The final issue is one of a good analytic technique for Piagetian data. In numerous places (Inhelder, 1956; Piaget, 1960; Pinard & Laurendeau, 1969), the major emphasis in investigating forms of thought has been on the ordering of the forms of thought, a task that seems to be well suited for Rasch scaling analysis. In the two known studies where Rasch scaling analysis has been used to replicate the order of items/tasks from a previous study, the technique has produced remarkable results. The results also provide strong support for the sample independence aspect of Rasch scaling analysis as the results were produced with empirical data generated from theoretically-based items/tasks, not from a simulation procedure. More work like this needs to be done to verify the usefulness and claims of Rasch scaling analysis.

References

Bart, W. M. (1972). Construction and validation of formal reasoning instruments. Psychological Reports, 30, 663–670.

Bingham-Newman, A. M., & Hooper, F. H. (1975). The search for the Woozle circa 1975: Commentary on Brainerd's observation. American Educational Research Journal, 12, 379–387.

Blasi, A., Hoeffel, E. C. (1974). Adolescence and formal operations. Human Development, 17, 344–363.

Bond, T. G. (1989). An investigation of the scaling of Piagetian formal operations. In P. Adey, J. Bliss, J. Head, & M. Shayer (Eds.), Adolescent development and school science (pp. 334–340). Philadelphia: Falmer Press.

Bond, T. G. (1995a). Piaget and measurement I: The twain really do meet. Archives de Psychologie, 63, 71–87.

Bond, T. G. (1995b). Piaget and measurement II: Empirical validation of the Piagetian model. Archives de Psychologie, 63, 155–185.

Bond, T. G., & Bunting, E. (1995). Piaget and measurement III: Reassessing the méthode clinique. Archives de Psychologie, 63, 231–255.

Brainerd, C. J. (1973). Judgments and explanations as criteria for the presence of cognitive structures. Psychological Bulletin, 79, 172–179.

Brainerd, C. J. (1974). Postmortem on judgments, explanations, and Piagetian cognitive structures. Psychological Bulletin, 81, 70–71.

Brainerd, C. J. (1975a). Rejoinder to Bingham-Newman and Hooper. American Educational Research Journal, 12, 389–394.

Brainerd, C. J. (1975b). Structures-of-the-whole and elementary education. American Educational Research Journal, 12, 369–378.

Brainerd, C. J. (1977). Response criteria in concept development research. Child Development, 48, 360–366.

Brainerd, C. J. (1978). The stage question in cognitive-developmental theory. The Behavioral and Brain Sciences, 1, 173–213.

Chapman, M. (1988). Constructive evolution Origins and development of Piaget's thought. New York: Cambridge University Press.

Chandler, M., & Chapman, M. (Eds.). (1991). Criteria for competence Controversies in the conceptualization and assessment of children's abilities. Hillsdale, NJ: Lawrence Erlbaum Associates.

Claparède, E. (1955). Preface. In J. Piaget, The language and thought of the child (M. Gabain, Trans., 2nd ed., pp. 11–18). Cleveland: Meridian Books. (Original work published 1931)

Clark, H. H., & Schober, M. F. (1992). Asking questions and influencing answers. In J. M. Tanur (Ed.), Questions about questions: Inquiries into the cognitive bases of surveys. New York: Russell Sage.

Cornish, G., & Wines, R. (1980). ACER mathematics profile series: Number test. Hawthorn, Australia: Australian Council for Educational Research. (ERIC Document Reproduction Service No. ED 189 173)

Demetriou, A., Efklides, A., Papadaki, M., Papantoniou, G., & Economou, A. (1993). Structure and development of causal-experimental thought: From early adolescence to youth. Developmental Psychology, 29, 480–497.

Demetriou, A., Efklides, A., Platsidou, M. (1993). The architecture and dynamics of developing mind: Experiential structuralism as a frame for unifying cognitive developmental theories. Monographs of the Society for Research in Child Development, 58(5–6, Serial 234).

Docherty, E. M. (1973). A test of Piaget's qualitative developmental hypothesis using the Rasch model. Dissertation Abstracts International, 34(3-B), 1255-1256.

Elliott, C. D. (1982). The development of the British Ability Scales (BAS). Paper presented at the 90th Annual Meeting of the American Psychological Association, Washington, DC. (ERIC Document Reproduction Service No. ED 227 130)

Farrell, M. A. (1969). The formal style: A review of the research. Journal of Research and Development in Education, 3, 111–118.

Flavell, J. H. (1963). The developmental psychology of Jean Piaget. Princeton, NJ: D. Van Nostrand.

Flavell, J. H., & Wohlwill, J. F. (1969). Formal and functional aspects of cognitive development. In D. Elkind & J. H. Flavell, Studies in cognitive development Essays in honor of Jean Piaget (pp. 67–120). New York: Oxford University Press.

Fox, C. M., & Gray, W. M. (1997, March). A theoretical and empirical investigation of misfitting persons and items based on cognitive development. In T. Bond (Chair), The measurement of psychological development. Symposium conducted at the Ninth International Objective Measurement Workshop, Chicago.

Gray, W. M. (1976)a. How Is Your Logic?. (Experimental edition, Form A). Boulder, CO: Biological Sciences Curriculum Study.

Gray, W. M. (1976b). How Is Your Logic?. (Experimental edition, Form B). Boulder, CO: Biological Sciences Curriculum Study.

Gray, W. M. (1990). Formal operational thought. In W. F. Overton (Ed.), Reasoning, necessity, and logic: Developmental perspectives (pp. 227–253). Hillsdale, NJ: Lawrence Erlbaum.

Gray, W. M., Bond, T. G., Fox, C. M., & Hofmann, R. J. (1996, June). A comparison of Rasch scaling, image analysis, and multiple hierarchical analysis of a set of Piagetian-based written problems representing different forms of thought and different logical operations. Poster session presented at the Twenty-sixth Annual Symposium of The Jean Piaget Society, Philadelphia.

Gray, W. M., & Fox, C. M. (1996, June). Rasch scaling of a set of Piagetian-based written problems representing different forms of thought and different logical operations. In T. L. Dawson (Chair), Rasch analysis in context. Symposium conducted at the Twenty-sixth Annual Symposium of The Jean Piaget Society, Philadelphia.

Hautamäki, J. (1989). The application of a Rasch model on Piagetian measures of stages of thinking. In P. Adey, J. Bliss, J. Head, & M. Shayer (Eds.), Adolescent development and school science (pp. 342–349). Philadelphia: Falmer Press.

Hubbs-Tait, L. (1986). The development of functionally similar and dissimilar operations of exclusion. Child Development 57, 934–941.

Inhelder, B. (1956). Criteria of the stages of mental development. In J. M. Tanner & B. Inhelder (Eds.), Discussions on child development A consideration of the biological, psychological, and cultural approaches to the understanding of human development and behaviour: Vol. 1. The proceedings of the first meeting of the World Health Organization Study Group on the Psychobiological Development of the Child Geneva 1953 (pp. 75–86). New York: International Universities Press.

Inhelder, B. (1971). Developmental theory and diagnostic procedures. In D. R. Green, M. P. Ford, G. B. Flamer (Eds.), Measurement and Piaget Proceedings of the CTB/McGraw-Hill conference on ordinal scales of cognitive development (pp. 148–171). New York: McGraw-Hill.

Inhelder, B., & Piaget, J. (1958). The growth of logical thinking from childhood to adolescence An essay on the construction of formal operational structures (A. Parson & S. Milgram, Trans.). New York: Basic Books. (Original work published 1955)

Inhelder, B., & Piaget, J. (1969). The early growth of logic in the child Classification and seriation (E. A. Lunzer & D. Papert, Trans.). New York: Norton Library. (Original work published 1959)

Kessen, W. (1996). American psychology just before Piaget. Psychological Science, 7, 196–199.

Kuhn, D., & Angelev, J. (1976). An experimental study of the development of formal operational thought. Child Development, 47, 697–706.

Kuhn, D., & Brannock, J. (1977). Development of the isolation of variables scheme in experimental and "natural experiment" contexts. Developmental Psychology, 13, 9–14.

Kuhn, D., & Ho, V. (1977). The development of schemes for recognizing additive and alternative effects in a "natural experiment" context. Developmental Psychology, 13, 515–516.

Kuhn, D., Langer, J., Kohlberg, & Haan, N. S. (1977). The development of formal operations in logical and moral judgment. Genetic Psychology Monographs, 95, 97–188.

Kuhn, D., & Phelps, E. (1979). A methodology for observing development of a formal reasoning strategy. In W. Damon (Editor-in-Chief) & D. Kuhn (Guest Ed.), New directions for child development: No. 5. Intellectual development beyond childhood (pp. 45–57). San Francisco: Jossey-Bass.

Lawson, A. E. (1985). A review of research on formal reasoning and science teaching. Journal of Research in Science Teaching, 22, 569–617.

Linacre, J. M., & Wright, B. D. (1991–1994). A user's guide to BIGSTEPS. Chicago: MESA Press.

Lopez, W. A. (1996). The resolution of ambiguity: An example from reading instruction. Unpublished doctoral dissertation, University of Chicago, Chicago.

Martorano, S. C. (1975, April). Formal operations thinking: Now you see it, now you don/t. Paper presented at the meeting of the Society for Research in Child Development, Denver.

Martorano, S. C. (1977). A developmental analysis of performance on Piaget's formal operations tasks. Developmental Psychology, 13, 666–672.

Masters, G. N. (1982). A Rasch model for partial credit scoring. Psychometrika, 47, 149–174.

Matthews, P. (1989). Discussion. (In discussion notes of The application of a Rasch model on Piagetian measures of stages of thinking by J. Hautamäki). In P. Adey, J. Bliss, J. Head, & M. Shayer (Eds.), Adolescent development and school science (p. 350). Philadelphia: Falmer Press.

Mueller, U., Reene, K., & Overton, W. F. (1994, June). Rasch-scaling of a deductive reasoning task. In Rasch-scaling. Symposium conducted at the twenty-fourth annual symposium of the Jean Piaget Society, Chicago.

Mueller, U., Winn, M., Overton, W. F. (1995, June). Rasch analysis of two recursive thinking tasks. Paper presented at the twenty-fifth annual symposium of the Jean Piaget Society, Berkeley, CA.

Neimark, E. D. (1975). Intellectual development during adolescence. In F. D. Horowitz (Ed.), Review of child development research (Vol. 4, pp. 541–594). Chicago: University of Chicago Press.

Neimark, E. D. (1979). Current status of formal operations research. Human Development, 22, 60–67.

Neimark, E. D. (1982). Adolescent thought: Transition to formal operations. In B. Wolman (Ed.), Handbook of developmental psychology (pp. 486–502). Englewood Cliffs, NJ: Prentice-Hall.

Noelting, G. (1976). Stages and mechanisms in the development of the concept of proportion in the child and adolescent (a preliminary report). In M. K. Poulsen, J. F. Magary, G. I. Lubin (Eds.), Piagetian theory and its implications for the helping professions Proceedings Fifth Interdisciplinary Seminar: Co-sponsored by University Affiliated Program, Childrens [sic] Hospital of Los Angeles and the University of Southern California School of Education (pp. 302–311). Los Angeles: University of Southern California.

Noelting, G., Coudé, G., & Rousseau, J. -P. (1995, June). Rasch analysis applied to multiple-domain tasks. Paper presented at the twenty-fifth annual symposium of the Jean Piaget Society, Berkeley, CA.

Noelting, G., Rousseau, J. -P., & Coudé, G. (1994, June). Rasch analysis applied to Piagetian-type problems. In Rasch analysis applied to Piagetian theory: Rethinking statistical methodology. Symposium conducted at the twenty-fourth annual symposium of the Jean Piaget Society, Chicago.

Parsons, A. (1958). Translators' introduction: A guide for psychologists. In B. Inhelder & J. Piaget, The growth of logical thinking from childhood to adolescence An essay on the construction of formal operational structures (A. Parson & S. Milgram, Trans., pp. vii–xx). New York: Basic Books. (Original work published 1955)

Piaget, J. (1955), The language and thought of the child (M. Gabain, Trans., 2nd ed.). Cleveland: Meridian Books. (Original work published 1931)

Piaget, J. (1956). Comments (In discussion of Criteria of the stages of mental development by B. Inhelder). In J. M. Tanner & B. Inhelder (Eds.), Discussions on child development A consideration of the biological, psychological, and cultural approaches to the understanding of human development and behaviour: Vol. 1. The proceedings of the first meeting of the World Health Organization Study Group on the Psychobiological Development of the Child Geneva 1953 (pp. 86–107). New York: International Universities Press.

Piaget, J. (1957). Logic and psychology (W. Mays & F. Whitehead, Trans.). New York: Basic Books. (Original work published 1953)

Piaget, J. (1960). The general problems of the psychobiological development of the child. In J. Tanner & B. Inhelder (Eds.), Discussions on child development A consideration of the biological, psychological, and cultural approaches to the understanding of human development and behaviour: Vol. 4. The proceedings of the fourth meeting of the World Health Organization Study Group on the Psychobiological

Development of the Child Geneva 1956 (pp. 3–27). New York: International Universities Press.

Piaget, J. (1965). The child's conception of number (C. Gattegno & F. M. Hodgson, Trans.). New York: Norton Library. (Original work published 1941)

Piaget, J. (1966). The psychology of intelligence (M. Piercy & D. E. Berlyne, Trans.). Totowa, NJ: Littlefield, Adams. (Original work published 1947)

Piaget, J., & Inhelder, B. (1974). The child's construction of quantities Conservation and atomism (A. J. Pomerans, Trans.). London: Routledge & Kegan Paul. (Original work published 1941)

Piaget, J., & Inhelder, B. (1975). The origin of the idea of chance in children (L. Leake, Jr., P. Burrell, & H. D. Fishbein, Trans.). New York: W. W. Norton. (Original work published 1951)

Pinard, a., & Laurendeau, M. (1969). "Stage" in Piaget's cognitive-developmental theory: Exegesis of a concept. In D. Elkind & J. H. Flavell, Studies in cognitive development Essays in honor of Jean Piaget (pp. 121–170). New York: Oxford University Press.

Rasch, G. (1980). Probabilistic models for some intelligence and attainment test. Chicago: University of Chicago Press.

Reese, H. W., & Schack, M. L. (1974). Comment on Brainerd's criteria for cognitive structures. Psychological Bulletin, 81, 67–69.

Shayer, M., Adey, P., & Wylam, H. (1981). Group tests of cognitive development ideals and a realization. Journal of Research in Science Teaching, 18, 157–168.

Shayer, M., Küchemann, D. E., & Wylam, H. (1976). The distribution of Piagetian stages of thinking in British middle and secondary school children. British Journal of Educational Psychology, 46, 164–173.

Snyder, S., & Sheehan, R. (1992). The Rasch measurement model: An introduction. Journal of Early Intervention, 16, 87–95.

Wilson, M. R. (1984). A psychometric model of hierarchical development. Dissertation Abstracts International, 45(4-A), 1079. Abstract from: OhioLINK File: Dissertation Abstracts Item: UMI# AAI0554227

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

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

Table 1

Distribution of Operations Across Items on Both Forms of How Is Your Logic?

| Concrete operation | Item | Formal operation | Item |
|---|---|---|---|
| Multiplication of classes | | Exclusion | |
| One-to-one correspondence | A7 | Make correct inclusion: Judgment | A2 |
| | B2 | | B9 |
| Many-to-one correspondence | A11 | Make correct inclusion: Explanation | A3 |
| | B13 | | B10 |
| | | Deny incorrect inclusions: Judgment | A9 |
| | | | B3 |
| | | Deny incorrect inclusions: | A10 |
| | | Explanation | B4 |
| Addition of asymmetrical relations | | Combinatorial thought | |
| Increasing series | A1 | Combinations | A5 |
| | B1 | | B6 |
| Decreasing series | A4 | Permutations | A6 |
| | B5 | | B7 |
| Multiplication of Relations[a] | | Proportional reasoning | |
| Increasing series-decreasing series | A8 | Formal I[b] | A12 |
| Decreasing series-decreasing series | B8 | Formal II[b] | B11 |

(table continues)

BEST COPY AVAILABLE

| Concrete operation | Item | Formal operation | Item |
|---|---|---|---|
| | | Explanation | A13 |
| | | | B12 |

[a]After using both forms of How Is Your Logic? for several studies (including this study) and finding the difficulty of the multiplication of relations items was much greater than expected given their design as concrete operational items, a number of participants were interviewed regarding their approaches to the problems. From these interviews, it was determined that the directions for items A8 and B8 were very ambiguous and a major cause of problems for participants because they did not understand the intent of the question given the graphics which were the focus of each problem. When the directions were elaborated orally, the items became much easier. Because all subjects in several studies, one of which occurred prior to this study, were not able to be interviewed orally about items A8 and B8, because How Is Your Logic? was designed to be a written instrument, not an orally administered instrument, because no other item on How Is Your Logic? had item difficulties that were unreasonable given their underlying form of thought, and because no other item elicited a multitude of questions about what to do, the data for items A8 and B8 are not reported. The non orally generated data for items A8 and B8 are available for those who wish to view them. The items are included in this table only for a sense of completeness regarding the composition of How Is Your Logic? and what participants were asked to do.

[b]Placement of numerical proportions into specific levels of formal operations suggested by Noelting (1976, personal communication, September 8, 1975).

Table 2

Average Category Measures and Category Fit for Two Best Stage Scoring Models

| | | 11233456 | | | 11234567 | | |
|---|---|---|---|---|---|---|---|
| | | | Infit | Outfit | | Infit | Outfit |
| Item | Category | Measure | MS | MS | Measure | MS | MS |
| A1 | 1 | .18 | 1.01 | .96 | .27 | 1.01 | .92 |
| | 2 | .44 | .98 | 1.08 | .49 | .99 | 1.04 |
| | 3 | .67 | 1.03 | 1.00 | .66 | 1.04 | 1.00 |
| A2 | 1 | .27 | 1.05 | 1.05 | .33 | 1.08 | 1.05 |
| | 2 | | | | | | |
| | 3 | .41 | .54 | .43 | .39 | .51 | .41 |
| | 4 | .53 | 1.26 | .92 | .63 | .65 | .51 |
| | 5 | .97 | .71 | .93 | .57 | 1.41 | 1.18 |
| | 6 | | | | .92 | .70 | .96 |
| A3 | 1 | .37 | .97 | .95 | .41 | .98 | .97 |
| | 2 | | | | | | |
| | 3 | .54 | 1.13 | 1.60 | .52 | 1.20 | 2.73 |
| | 4 | .57 | 1.70 | 1.32 | .55 | 1.25 | .86 |
| | 5 | 1.03 | .68 | .63 | .59 | 1.82 | 1.41 |
| | 6 | | | | .97 | .66 | .61 |

(table continues)

| Item | Category | 11233456 Measure | 11233456 Infit MS | 11233456 Outfit MS | 11234567 Measure | 11234567 Infit MS | 11234567 Outfit MS |
|---|---|---|---|---|---|---|---|
| A4 | 1 | .06 | .98 | .83 | .18 | .99 | .84 |
| | 2 | .49 | 1.01 | 1.10 | .53 | 1.01 | 1.08 |
| | 3 | .68 | 1.06 | 1.01 | .67 | 1.05 | 1.01 |
| A5 | 1 | -.15 | .87 | .89 | -.06 | .92 | .92 |
| | 2 | | | | | | |
| | 3 | .49 | 1.02 | 1.01 | .33 | .82 | .80 |
| | 4 | | | | .58 | 1.09 | 1.07 |
| | 5 | .86 | .97 | .95 | | | |
| | 6 | | | | .82 | .96 | .96 |
| A6 | 1 | .29 | .97 | 1.00 | .34 | 1.00 | 1.05 |
| | 2 | | | | | | |
| | 3 | .58 | .82 | .77 | .47 | .75 | .72 |
| | 4 | | | | .63 | .79 | .68 |
| | 5 | .94 | .90 | .88 | | | |
| | 6 | 1.29 | .77 | .83 | .88 | .90 | .89 |
| | 7 | | | | 1.18 | .75 | .84 |

(table continues)

| | | 11233456 | | | 11234567 | | |
|---|---|---|---|---|---|---|---|
| Item | Category | Measure | Infit MS | Outfit MS | Measure | Infit MS | Outfit MS |
| A7 | 1 | -.01 | .88 | .65 | .08 | .88 | .66 |
| | 2 | .48 | .98 | 1.06 | .51 | .98 | 1.03 |
| | 3 | .72 | .88 | .95 | .70 | .89 | .95 |
| A9 | 1 | .50 | 1.53 | 1.57 | .53 | 1.56 | 1.69 |
| | 2 | | | | | | |
| | 3 | .51 | .96 | 1.11 | .41 | .59 | .86 |
| | 4 | .63 | 1.19 | 1.26 | .76 | .84 | .81 |
| | 5 | | | | .65 | 1.29 | 1.39 |
| | 6 | 1.20 | .80 | .77 | | | |
| | 7 | | | | 1.10 | .79 | .76 |
| A10 | 1 | .47 | 1.15 | 1.09 | .49 | 1.16 | 1.12 |
| | 2 | | | | | | |
| | 3 | .72 | 1.16 | 1.05 | .63 | 1.32 | 1.20 |
| | 4 | .70 | 1.32 | 1.43 | .76 | 1.03 | .86 |
| | 5 | | | | .70 | 1.38 | 1.50 |
| | 6 | 1.29 | .81 | .70 | | | |
| | 7 | | | | 1.18 | .79 | .68 |

(table continues)

| | | 11233456 | | | 11234567 | | |
|---|---|---|---|---|---|---|---|
| Item | Category | Measure | Infit MS | Outfit MS | Measure | Infit MS | Outfit MS |
| A11 | 1 | .05 | .91 | .73 | .17 | .93 | .75 |
| | 2 | .42 | .96 | .87 | .45 | .96 | .85 |
| | 3 | .71 | .96 | .98 | .69 | .97 | .99 |
| A12 | 1 | .45 | 1.09 | 1.10 | .48 | 1.08 | 1.08 |
| | 2 | | | | | | |
| | 3 | .69 | .95 | 1.17 | .67 | .90 | 1.26 |
| | 4 | | | | | | |
| | 5 | .99 | 1.09 | 1.36 | | | |
| | 6 | | | | .94 | 1.12 | 1.49 |
| A13 | 1 | .48 | .95 | .97 | .50 | .96 | .98 |
| | 2 | | | | | | |
| | 3 | .66 | 1.10 | 1.39 | .71 | .67 | .66 |
| | 4 | .98 | .78 | .70 | .58 | 1.58 | 2.56 |
| | 5 | 1.13 | .90 | .83 | .92 | .83 | .74 |
| | 6 | 1.58 | .76 | .64 | 1.05 | .90 | .83 |
| | 7 | | | | 1.46 | .73 | .61 |

(table continues)

| | | 11233456 | | | 11234567 | | |
|---|---|---|---|---|---|---|---|
| Item | Category | Measure | Infit MS | Outfit MS | Measure | Infit MS | Outfit MS |
| B1 | 1 | .11 | 1.01 | .84 | .25 | 1.01 | .87 |
| | 2 | .44 | .99 | 1.02 | .47 | .99 | .98 |
| | 3 | .67 | 1.03 | 1.00 | .66 | 1.03 | 1.00 |
| B2 | 1 | .10 | .94 | .73 | .19 | .94 | .75 |
| | 2 | .33 | .89 | .71 | .38 | .91 | .73 |
| | 3 | .72 | .89 | .95 | .70 | .90 | .95 |
| B3 | 1 | .46 | 1.47 | 1.45 | .49 | 1.51 | 1.52 |
| | 2 | | | | | | |
| | 3 | .52 | .84 | .92 | .43 | .59 | .68 |
| | 4 | .60 | 1.27 | 1.22 | .75 | .65 | .84 |
| | 5 | | | | .62 | 1.39 | 1.28 |
| | 6 | 1.17 | .80 | .77 | | | |
| | 7 | | | | 1.08 | .77 | .76 |

(table continues)

| | | 11233456 | | | 11234567 | | |
|---|---|---|---|---|---|---|---|
| Item | Category | Measure | Infit MS | Outfit MS | Measure | Infit MS | Outfit MS |
| B4 | 1 | .48 | 1.26 | 1.17 | .50 | 1.26 | 1.20 |
| | 2 | | | | | | |
| | 3 | .68 | 1.21 | 1.29 | .65 | 1.42 | 1.62 |
| | 4 | .65 | 1.48 | 1.40 | .69 | 1.20 | 1.20 |
| | 5 | | | | .65 | 1.60 | 1.58 |
| | 6 | 1.25 | .85 | .79 | | | |
| | 7 | | | | 1.15 | .83 | .77 |
| B5 | 1 | .57 | 1.07 | 1.43 | .58 | 1.05 | 1.30 |
| | 2 | .69 | 1.06 | 1.31 | .69 | 1.05 | 1.24 |
| | 3 | .65 | 1.13 | 1.03 | .65 | 1.10 | 1.03 |
| B6 | 1 | .08 | 1.03 | .98 | .12 | 1.07 | 1.08 |
| | 2 | | | | | | |
| | 3 | .52 | 1.03 | 1.07 | .36 | .86 | .88 |
| | 4 | | | | .59 | 1.06 | 1.13 |
| | 5 | .87 | 1.00 | .99 | | | |
| | 6 | | | | .82 | .98 | .98 |

(table continues)

| | | 11233456 | | | 11234567 | | |
|---|---|---|---|---|---|---|---|
| Item | Category | Measure | Infit MS | Outfit MS | Measure | Infit MS | Outfit MS |
| B7 | 1 | .39 | .99 | .98 | .43 | 1.01 | 1.00 |
| | 2 | | | | | | |
| | 3 | .72 | .84 | .73 | .64 | .87 | .85 |
| | 4 | | | | .74 | .79 | .59 |
| | 5 | .97 | 1.13 | 1.08 | | | |
| | 6 | 1.12 | .99 | 1.11 | .92 | 1.15 | 1.09 |
| | 7 | | | | 1.03 | .97 | 1.12 |
| B9 | 1 | .31 | 1.11 | 1.11 | .36 | 1.14 | 1.11 |
| | 2 | | | | | | |
| | 3 | .40 | .59 | .55 | .37 | .44 | .33 |
| | 4 | .54 | 1.32 | 1.04 | .83 | 1.15 | 1.68 |
| | 5 | .94 | .76 | .81 | .57 | 1.40 | 1.06 |
| | 6 | | | | .90 | .74 | .78 |
| B10 | 1 | .39 | .96 | .91 | .43 | .96 | .91 |
| | 2 | | | | | | |
| | 3 | .57 | 1.07 | .73 | .51 | .77 | .55 |
| | 4 | .61 | 1.46 | 1.19 | .64 | 1.22 | .82 |
| | 5 | 1.04 | .77 | .69 | .63 | 1.54 | 1.27 |
| | 6 | | | | .98 | .76 | .65 |

(table continues)

| | | 11233456 | | | 11234567 | | |
|---|---|---|---|---|---|---|---|
| | | | Infit | Outfit | | Infit | Outfit |
| Item | Category | Measure | MS | MS | Measure | MS | MS |
| B11 | 1 | .50 | 1.47 | 2.23 | .52 | 1.44 | 2.19 |
| | 2 | | | | | | |
| | 3 | .47 | 1.11 | 2.52 | .48 | .93 | 1.98 |
| | 4 | | | | | | |
| | 5 | | | | | | |
| | 6 | .95 | 1.40 | 2.16 | | | |
| | 7 | | | | .90 | 1.40 | 2.26 |
| B12 | 1 | .46 | .96 | .97 | .48 | .97 | .98 |
| | 2 | | | | | | |
| | 3 | .58 | 1.30 | 1.70 | .60 | 1.04 | 1.27 |
| | 4 | .92 | .81 | .68 | .57 | 1.62 | 2.38 |
| | 5 | 1.14 | .81 | .72 | .88 | .86 | .72 |
| | 6 | 1.50 | .85 | .75 | 1.06 | .81 | .71 |
| | 7 | | | | 1.39 | .80 | .72 |
| B13 | 1 | -.05 | .93 | .75 | .05 | .93 | .74 |
| | 2 | .27 | .96 | .63 | .33 | .96 | .67 |
| | 3 | .68 | .82 | .96 | .67 | .83 | .96 |

Table 3

Rasch Item Statistics in Measure Order for 11233456 Stage Scoring Model

| | | | | Infit | | Outfit | | |
|---|---|---|---|---|---|---|---|---|
| Operation | Item | Measure | Error | MS | zstd | MS | zstd | $r_{pb}$ |
| Proportional Reasoning: | A13 | 1.42 | .03 | .89 | -1.8 | .90 | -1.0 | .52 |
| Explanation | B12 | 1.38 | .03 | .90 | -1.8 | .89 | -1.2 | .52 |
| Deny incorrect inclusions: | A10 | 1.08 | .03 | 1.04 | .6 | 1.03 | .3 | .46 |
| Explanation | | | | | | | | |
| Combinatorial thought: | B7 | 1.05 | .03 | 1.01 | .3 | 1.00 | .0 | .48 |
| Permutations | | | | | | | | |
| Deny incorrect inclusions: | B4 | 1.04 | .03 | 1.11 | 1.9 | 1.09 | .9 | .41 |
| Explanation | | | | | | | | |
| Combinatorial thought: | A6 | .93 | .03 | .90 | -2.1 | .90 | -1.8 | .52 |
| Permutations | | | | | | | | |
| Proportional reasoning: | A12 | .85 | .03 | 1.08 | 1.6 | 1.24 | 2.2 | .42 |
| Judgment | B11 | .83 | .02 | 1.42 | 6.6 | 2.21 | 2.9 | .31 |
| Deny incorrect inclusions: | A9 | .78 | .03 | 1.15 | 2.7 | 1.19 | .28 | .36 |
| Judgment | | | | | | | | |

(table continues)

| | | | | Infit | | Outfit | | |
|---|---|---|---|---|---|---|---|---|
| Operation | Item | Measure | Error | MS | zstd | MS | zstd | $r_{pb}$ |
| Make correct inclusion: Explanation | B10 | .76 | .03 | .89 | -2.4 | .81 | -2.1 | .55 |
| Deny incorrect inclusions: Judgment | B3 | .73 | .03 | 1.11 | .19 | 1.10 | 1.4 | .39 |
| Make correct inclusion: Explanation | A3 | .68 | .03 | .88 | -2.6 | .88 | -1.4 | .54 |
| Make correct inclusion: | B9 | .33 | .03 | .93 | -1.5 | .91 | -1.2 | .49 |
| Judgment | A2 | .30 | .03 | .87 | -2.4 | .89 | -1.5 | .53 |
| Combinatorial thought: | B6 | -.24 | .04 | 1.02 | .4 | 1.02 | .3 | .32 |
| Combinations | A5 | -.36 | .04 | .98 | -.5 | .97 | -.5 | .34 |
| Multiplication of classes | B2 | -.92 | .09 | .92 | -.7 | .75 | -1.8 | .30 |
| | A7 | -.93 | .08 | .91 | -.9 | .83 | -1.4 | .28 |
| | A11 | -1.00 | .09 | .93 | -.6 | .80 | -1.4 | .26 |
| | B13 | -1.49 | .13 | .93 | -.4 | .72 | -1.1 | .20 |

(table continues)

| | | | | Infit | | Outfit | | |
|---|---|---|---|---|---|---|---|---|
| Operation | Item | Measure | Error | MS | zstd | MS | zstd | $r_{pb}$ |
| Addition of asymmetrical | A1 | -1.71 | .14 | 1.00 | .0 | 1.02 | .1 | .12 |
| relations | A4 | -1.76 | .12 | 1.01 | .1 | 1.02 | .1 | .14 |
| | B5 | -1.87 | .14 | 1.07 | .4 | 1.33 | 1.5 | -.02 |
| | B1 | -1.88 | .15 | 1.00 | .0 | .96 | -.2 | .12 |

## Figure Captions

Figure 1. Item separation and person separation for nine best stage scoring models.

Figure 2. Item fit for nine best stage scoring models by outfit and infit.

Figure 3. Person fit for nine best stage scoring models by outfit and infit.

Figure 4. Map of persons and items on the same logit scale for six stage scoring model.

Figure 5. Item logit scores for Gray & Fox (1997) and Fox & Gray (1997) Six Category Stage Scoring Model (11233456).

Figure 6. Item logit scores for Gray & Fox (1997) and Fox & Gray (1997) Four Category Stage Scoring Model (11233444).

Figure 1. Item Separation and Person Separation for Nine Best Stage Scoring Models

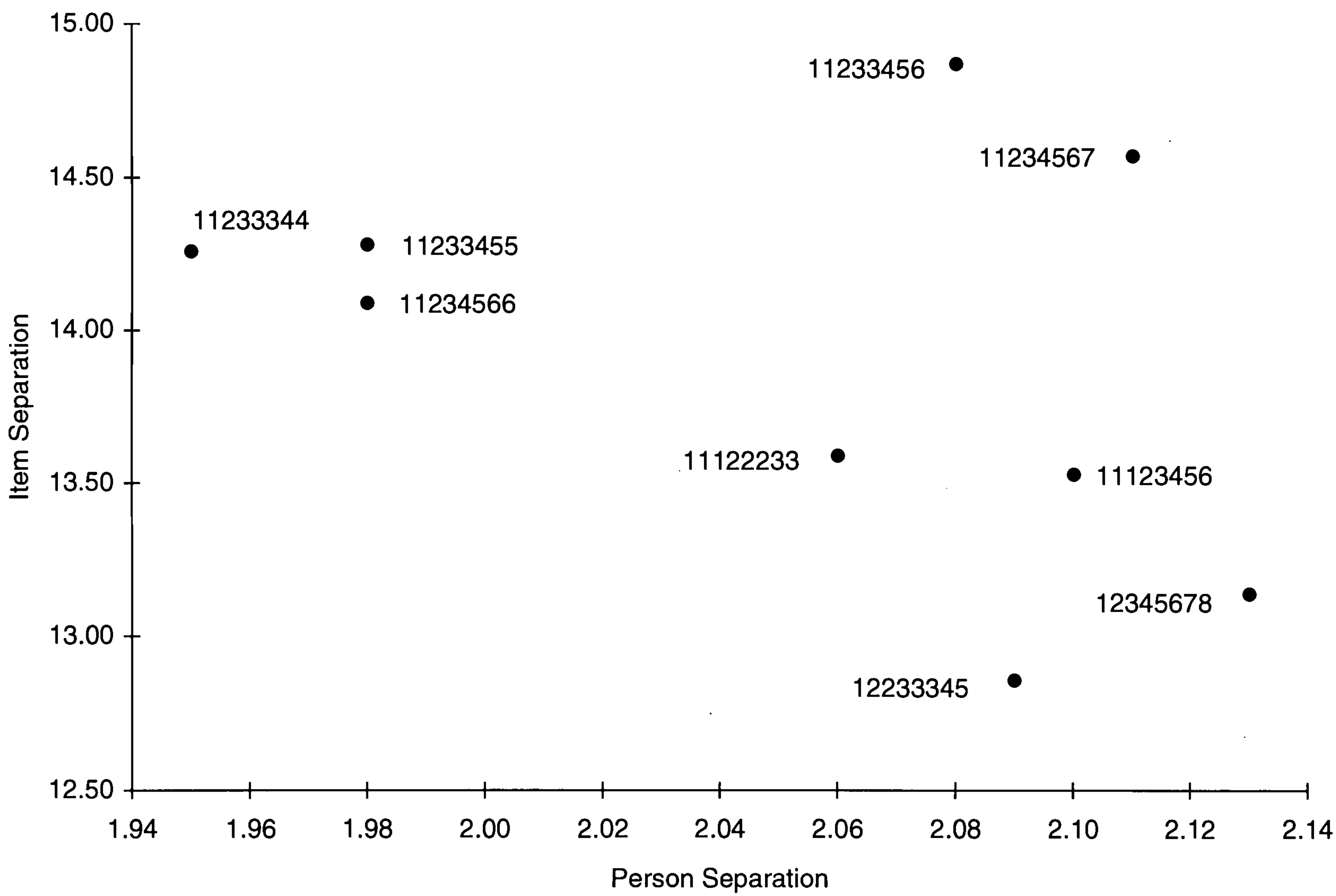


Figure 2. Item Fit for Nine Best Stage Scoring Models by Outfit and Infit

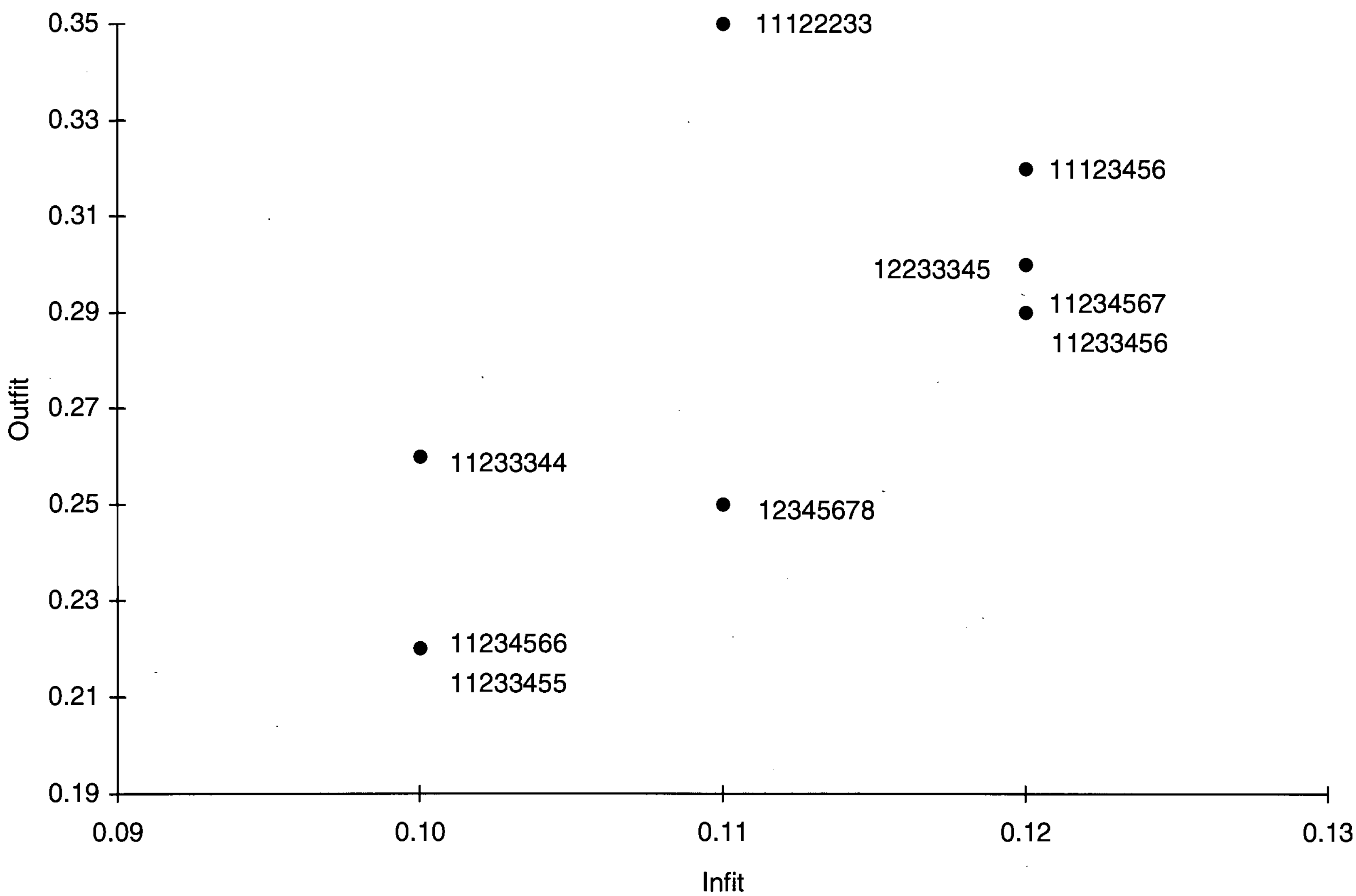


Figure 3. Person Fit for Nine Best Stage Scoring Models by Outfit and Infit

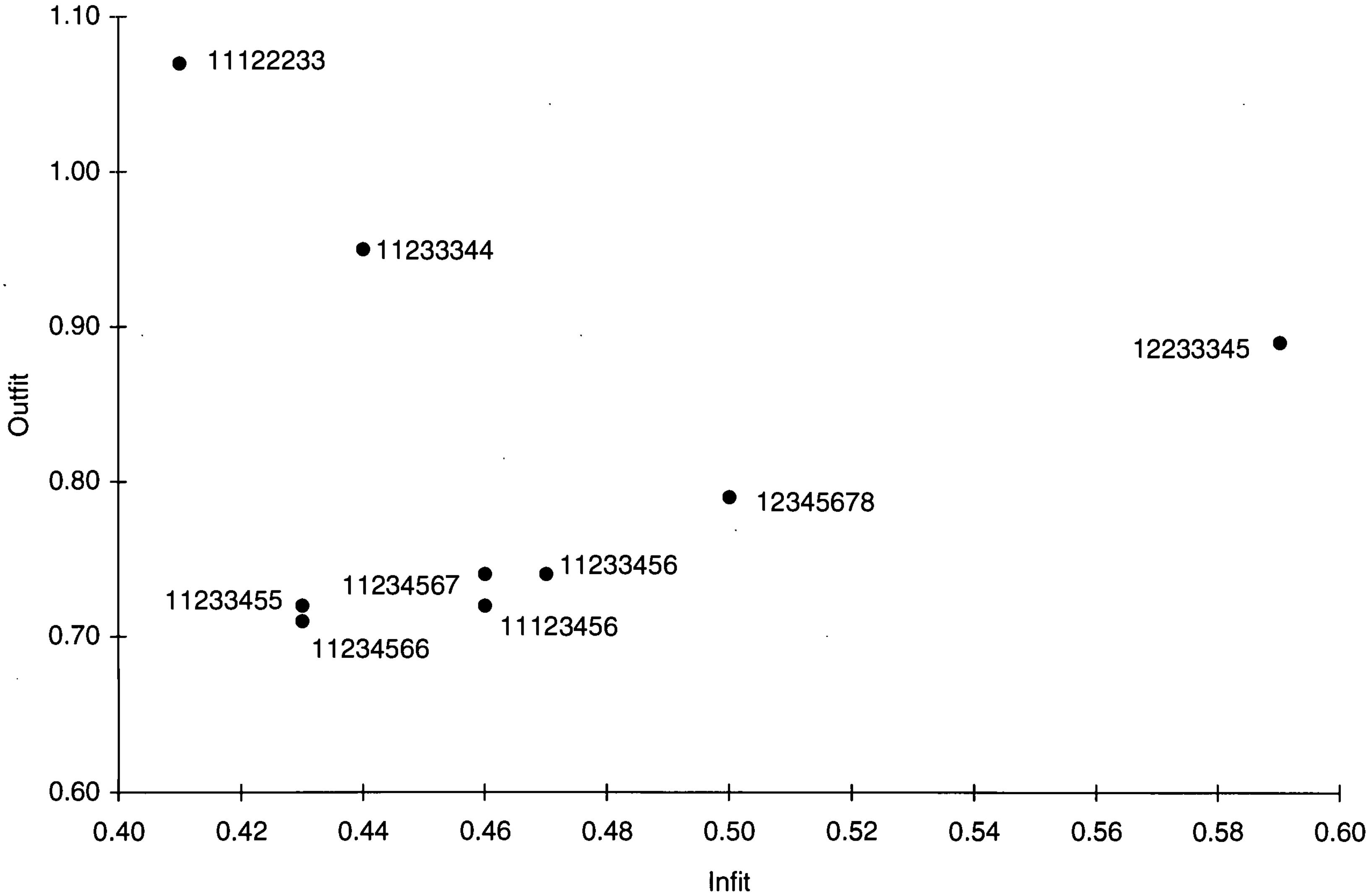


Figure 4. Map of Persons and Items on the Same Logit Scale for Six Category Stage Scoring Model

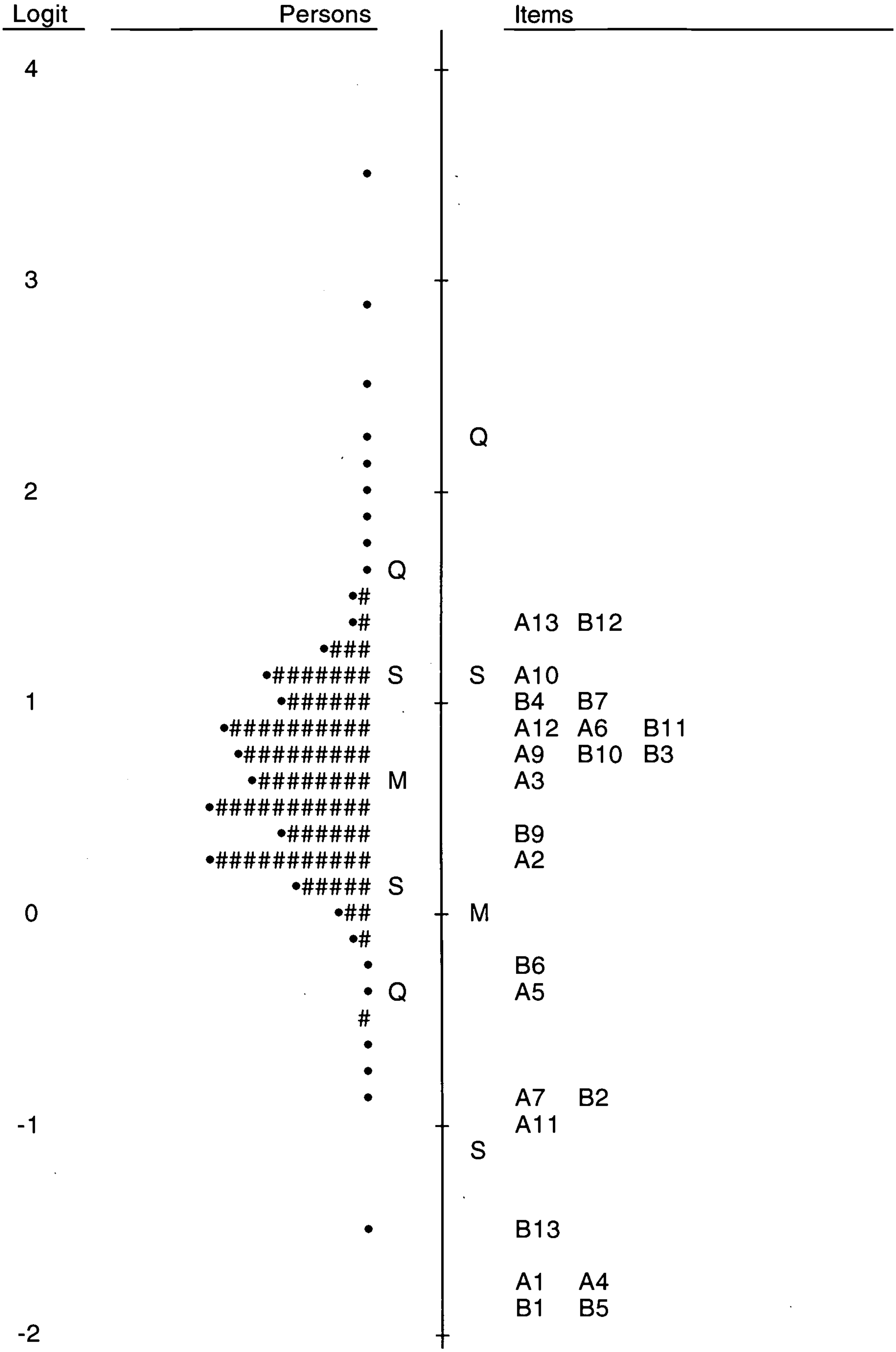


Figure 5. Item Logit Scores for Gray & Fox (1997) and Fox & Gray (1997) Six Category Stage Scoring Model (11233456)

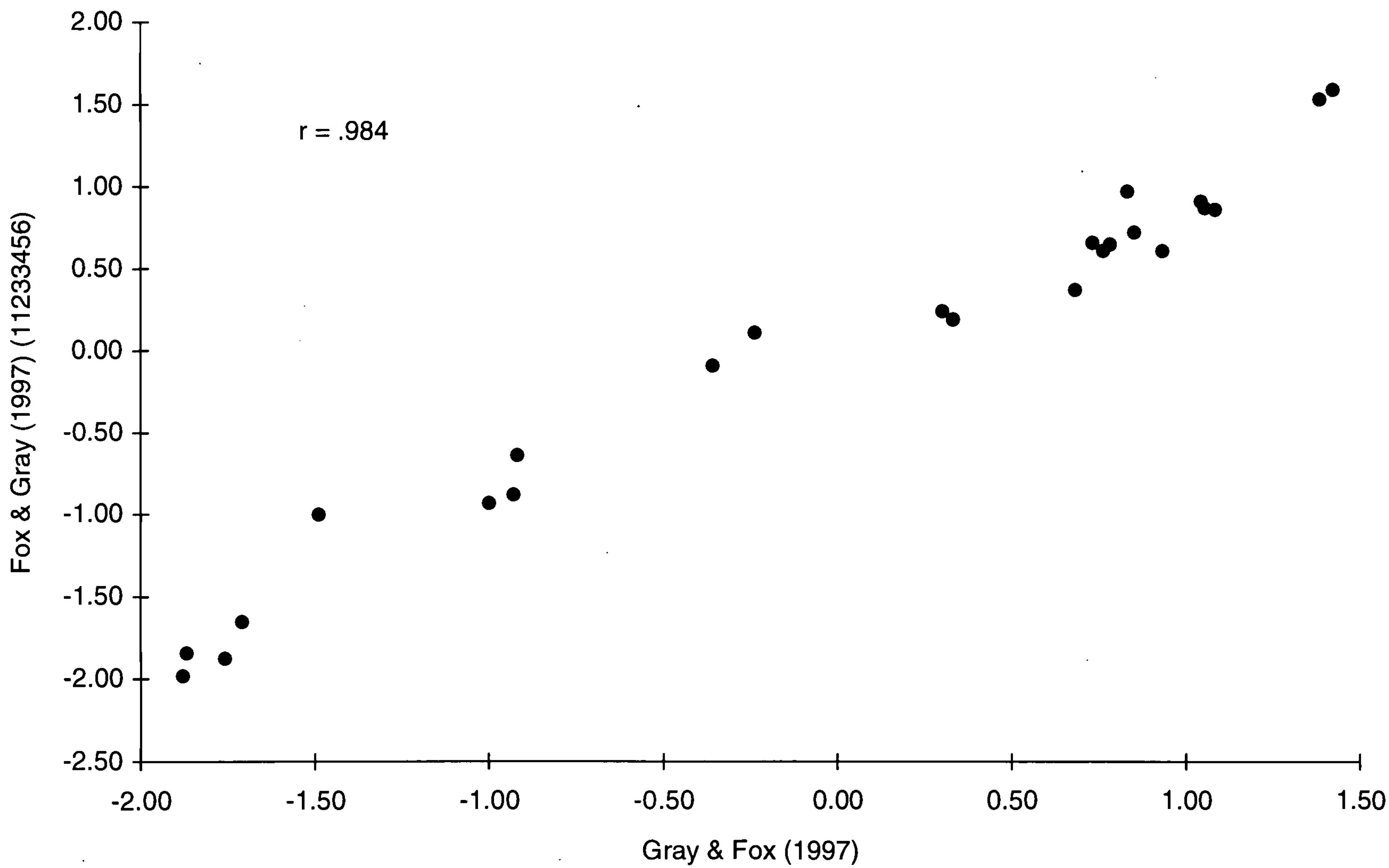
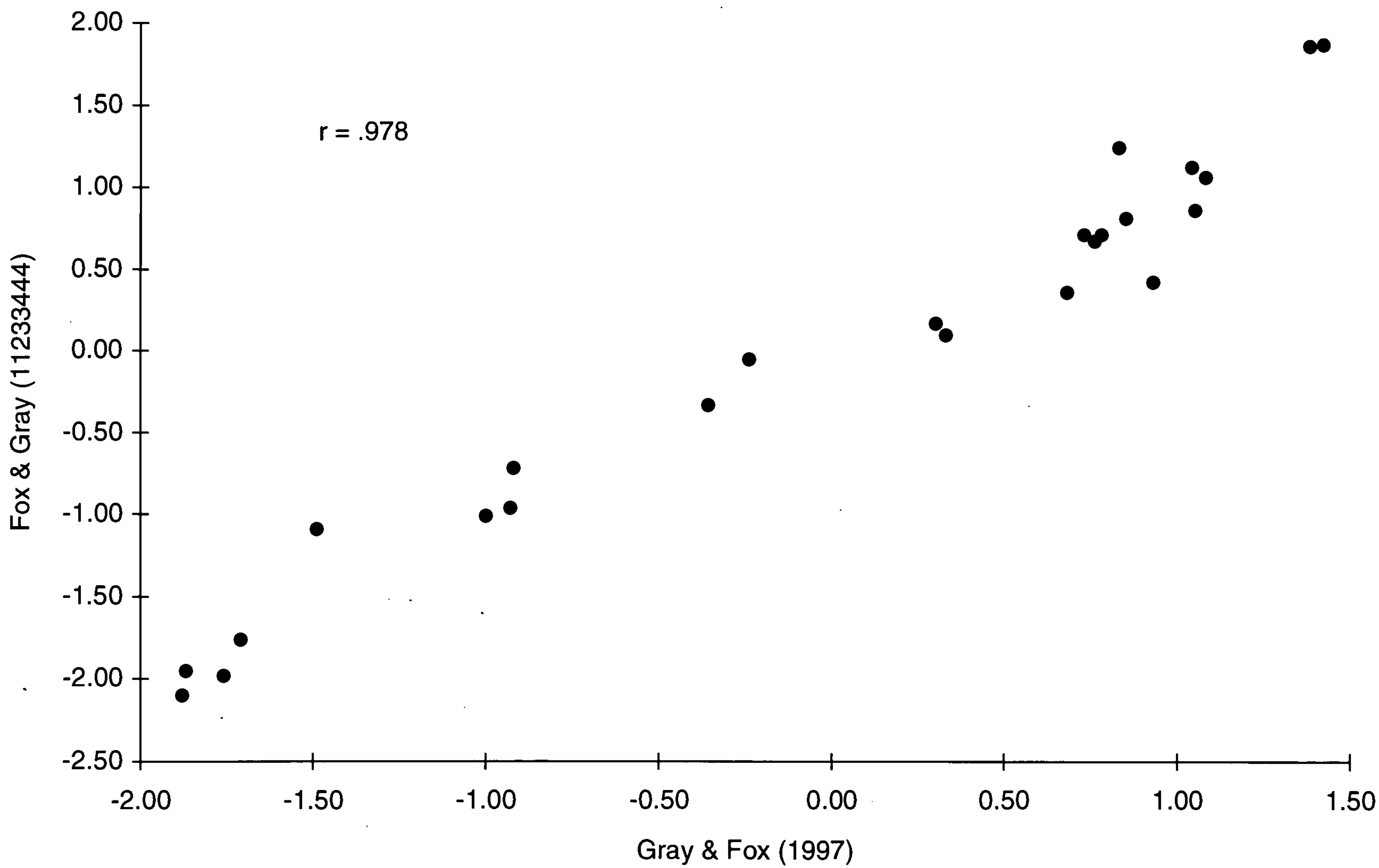


Figure 6. Item Logit Scores for Gray & Fox (1997) and Fox & Gray (1997) Four Category Stage Scoring Model (11233444)



AREA 1997

TM027307

**U.S. DEPARTMENT OF EDUCATION**
*Office of Educational Research and Improvement (OERI)*
*Educational Resources Information Center (ERIC)*

ERIC

# REPRODUCTION RELEASE

(Specific Document)

## I. DOCUMENT IDENTIFICATION:

| | |
|---|---|
| Title: RASCH SCALING OF DIFFERENT FORMS OF THOUGHT AND DIFFERENT LOGICAL OPERATIONS | |
| Author(s): WILLIAM M GRAY CHRISTINE M FOX | |
| Corporate Source: UNIVERSITY OF TOLEDO | Publication Date: 1997 |

## II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

[X] Sample sticker to be affixed to document

**Check here**
Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Sample
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

**Level 1**

Sample sticker to be affixed to document [ ]

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY
Sample
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

**Level 2**

**or here**
Permitting reproduction in other than paper copy.

## Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

| | |
|---|---|
| Signature: William M. Gray | Position: PROFESSOR OF EDUCATIONAL PSYCHOLOGY |
| Printed Name: WILLIAM M GRAY | Organization: UNIVERSITY OF TOLEDO |
| Address: EDUCATIONAL PSYCHOLOGY UNIVERSITY OF TOLEDO TOLEDO OH 43606-3390 | Telephone Number: (419) 530-2481 |
| | Date: 16 APRIL 97 |



OVER

CUA

THE CATHOLIC UNIVERSITY OF AMERICA
*Department of Education, O'Boyle Hall*
*Washington, DC 20064*
*202 319-5120*

February 21, 1997

Dear AERA Presenter,

Congratulations on being a presenter at AERA[1]. The ERIC Clearinghouse on Assessment and Evaluation invites you to contribute to the ERIC database by providing us with a printed copy of your presentation.

Abstracts of papers accepted by ERIC appear in *Resources in Education (RIE)* and are announced to over 5,000 organizations. The inclusion of your work makes it readily available to other researchers, provides a permanent archive, and enhances the quality of *RIE*. Abstracts of your contribution will be accessible through the printed and electronic versions of *RIE*. The paper will be available through the microfiche collections that are housed at libraries around the world and through the ERIC Document Reproduction Service.

We are gathering all the papers from the AERA Conference. We will route your paper to the appropriate clearinghouse. You will be notified if your paper meets ERIC's criteria for inclusion in *RIE*: contribution to education, timeliness, relevance, methodology, effectiveness of presentation, and reproduction quality. You can track our processing of your paper at http://ericae2.educ.cua.edu.

Please sign the Reproduction Release Form on the back of this letter and include it with **two** copies of your paper. The Release Form gives ERIC permission to make and distribute copies of your paper. It does not preclude you from publishing your work. You can drop off the copies of your paper and Reproduction Release Form at the **ERIC booth (523)** or mail to our attention at the address below. Please feel free to copy the form for future or additional submissions.

Mail to:

AERA 1997/ERIC Acquisitions
The Catholic University of America
O'Boyle Hall, Room 210
Washington, DC 20064

This year ERIC/AE is making a **Searchable Conference Program** available on the AERA web page (http://aera.net). Check it out!

Sincerely,

Lawrence M. Rudner, Ph.D.
Director, ERIC/AE

[1]If you are an AERA chair or discussant, please save this form for future use.