DOCUMENT RESUME

ED 373 086 TM 021 975

AUTHOR Bolt, Daniel; Ackerman, Terry
TITLE An Examination of the Influence of Expository and Narrative Passages on the Dimensionality of the IGAP Reading Test.
PUB DATE Apr 94
NOTE 30p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994). For related documents, see TM 021 976-977.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Achievement; Context Effect; Correlation; Educational Objectives; Elementary School Students; Elementary Secondary Education; Factor Analysis; Grade 3; Grade 6; Grade 8; Grade 10; *Item Response Theory; Reading Achievement; Reading Comprehension; *Reading Tests; Secondary School Students; *State Programs; Test Format; *Testing Programs; *Test Results
IDENTIFIERS *Dimensionality (Tests); Expository Text; Illinois; *Illinois Goal Assessment Program; Multidimensionality (Tests); Narrative Text

ABSTRACT

The 1993 Illinois Goal Assessment Program (IGAP) Reading Tests measured reading comprehension using both narrative and expository reading passages. Noticeable differences in mean scaled scores occurred depending on whether the 1993 results were equated back to the 1992 narrative test or the 1993 expository test (Hsu and Ackerman, 1994). In an attempt to explain these disparate results, this investigation examines whether combining the two types of passages on a reading test induces multidimensionality. The tests were administered to all Illinois students in grades 3, 6, 8, and 10, and tests from nearly 5,000 examinees were used in the analysis. Results from a principal components factor analysis, a statistical test of dimensionality, and multidimensional item-response theory suggest that items based on the narrative and expository passages measure distinct, yet highly correlated, dimensions. Four figures and five tables present study findings. (Contains 14 references.) (Author/SLD)

An Examination of the Influence of Expository and Narrative Passages on the Dimensionality of the IGAP Reading Test

Daniel Bolt
Terry Ackerman
University of Illinois at Urbana-Champaign

Paper presented at the 1994 AERA Annual Meeting, New Orleans

Running Head: READING TEST

Abstract

The 1993 Illinois Goal Assessment Program (IGAP) Reading Tests measured reading comprehension using both narrative and expository reading passages. Noticable differences in mean scaled score occurred depending on whether the 1993 results were equated back to the 1992 narrative test or the 1992 expository test (Hsu & Ackerman, 1994). In an attempt to explain these disparate results, this investigation examines whether combining the two types of passages on a reading test induces multidimensionality. Results from a principal components factor analysis, a statistical test of dimensionality, and multidimensional item response theory suggest that items based on the narrative and expository passages measure distinct, yet highly correlated dimensions.

An Examination of the Influence of Expository and Narrative Passages on the Dimensionality of the IGAP Reading Test

Introduction

In recent years, reading research has focused on the way textual features, response formats and examinee abilities interact in the assessment of reading comprehension. Reading is typically viewed as an interactive process in which the reader combines various sources of information with material presented in a text to construct meaning. These sources of information include general knowledge of the reading process and reading strategies as well as knowledge specific to the text, such as familiarity with the topic or the genre within which the information is presented. Because readers can be expected to differ in these areas, this representation of the reading process emphasizes the uniqueness of the meaning constructed from a passage by each reader. The purpose of this paper is to summarize an investigation about the extent to which the assessment process of reading comprehension skills required by the narrative and expository genres induce multidimensionality.

Background

The reading test of the Illinois Goal Assessment Program reflects a dynamic and complex view of reading in its innovative approach to measuring reading comprehension. This is demonstrated both by the nature of the passages and the

types of items used. The reading selections are complete passages of text and subsequently contain structures with which students are familiar, namely, those with a beginning, a middle, and an end. They concern topics that most students recognize and are similar to those that students are exposed to at school or at home. Because these passages are longer than those typically encountered on reading comprehension tests, a greater diversity of questions can be asked emphasizing a wider range of thinking skills. The test also simulates typical classroom situations in which there may be more than one correct inference to a question by employing a scoring format that enables students to receive partial credit for each correct and incorrect inference that they can identify (Illinois State Board of Education, 1991).

In 1993, the IGAP reading tests were administered to all Illinois students in grades 3, 6, 8, and 10. Each test contained two reading passages, one representing the narrative (story-type) genre and the other the expository (information-type) genre. Prior to reading each passage, students answered a pair of questions designed to assess familiarity with the topic of the reading passage. After completing the two items, students were instructed to read each passage and then respond to 15 items assessing their abilities to construct meaning from the passage. Each item consisted of a question followed by 5 conclusions, at least one of which, and no more than three of which, were correct. The scoring rubric graded each item on a zero to five scale (corresponding to number correct)

producing raw scores which could range from zero to 150 . Upon completion of the reading comprehension items, students were instructed to complete a final series of questions related to the passage concerning use of appropriate reading strategies. The outcome measure of the test is based on the examinee's performance on the 15 reading comprehension items. Each examinee's raw score was then converted to a standardized score for reporting.

Previously, equating of test results was performed using techniques from classical test theory. The potential employment of procedures based on item response theory (IRT) requires preliminary consideration of several issues, one of which concerns the assumption of unidimensionality. The possible violation of this assumption is suggested by the findings discussed by Hsu and Ackerman (1994) in which IRT true score equating with the IGAP reading test produced different results depending on whether linking was performed using items from previously administered narrative or expository passages. One likely explanation for these results is that questions based on the narrative passage are actually measuring a different skill (or different composite of multiple skills) than those based on the expository passage. This would be consistent with the description of the reading process discussed earlier in which knowledge of a particular text structure plays an important role in the process of comprehension. The objective of this paper is to investigate the conjecture that combining both types of passages into one test results in multidimensionality.

Differences between the narrative and expository genres have been the focus of a considerable amount of research in reading comprehension. Narrative passages typically involve familiar themes in which major characters engage in action towards resolution of some problem or achievement of some goal. Expository passages, on the other hand, attempt to present new information and can take a variety of forms, including description, comparison and contrast, and cause-and-effect (Leslie and Caldwell, 1988). Because narrative texts contain certain inherent structural characteristics, they possess greater predictabilty and thus allow readers to more readily "fill in gaps" left by missing information. These characteristics are absent, however, in expository texts, and without the knowledge and experience necessary for interpretation of less coherent text, readers are typically left to impose thier own organization on the content for the passage to achieve meaning (Beck & McKeown, 1989). As a result, it is not surprising that expository texts are considered more difficult. While there certainly exists a large amount of variability within each of these genres, these distinctions between the two are sufficient to warrant consideration of each when assessing children's reading abilities (Leslie & Caldwell, 1988).

The fact that students have greater difficulty in working with expository passages has been attributed to a lack of exposure to appropriate expository reading material (Beck and McKeown, 1989) and the need for knowledge concerning expository text structures (Armbruster, Anderson, & Ostertag, 1989;

Englert &Hiebert, 1989). This unfamiliarity with expository reading material is especially disturbing considering the large dependence on such reading material in all areas of schooling at the intermediate grade levels (Spiro & Taylor, 1980). Such concerns have led to investigations into the process by which awareness of expository text structure emerges (Englert & Hiebert, 1984) as well as specific methods by which abilities in this area can be strengthened, such as the teaching of text structure (Piccolo, 1987;Armbruster, et. al. 1989) and improvements in basal readers (Beck & McKeown, 1989). As schools apply such suggestions, it may become increasingly important that progress be monitored in terms of reading ability specific to the expository genre. Because a major purpose of the IGAP tests is to simultaneously provide schools with information for policy-making and chart progress due to policy changes, the issue of whether or not to report separate scale scores for the two types of passages because of multidimensionallity needs to be investigated.

Method.

Several procedures were used to examine the possibility that multidimensionality exists in the IGAP Reading tests due to the inclusion of both narrative and expository passages. Data were obtained from administrations of the 1993 IGAP Reading Tests at Grades 3, 6 and 8 with approximately 5000 examinees at each grade level. A principal components factor analysis was performed considering performance based on the partial credit scoring for each of

the thirty items. The computer programs DIMTEST (Stout, Douglas, Junker & Roussos, 1993) and NOHARM (Fraser & McDonald, 1988) were also used to investigate multidimensionality at the level of the individual responses to each question.

Results and Discussion

Factor Analysis

A two-factor solution was obtained by performing a principal components factor analysis on the thirty items at each grade level. Table 1 contains the eigenvalues obtained for each grade and appears to suggest that the reading test is close to being unidimensional in all cases. The factor patterns following a promax rotation are displayed in Table 2 . Items 1 through 15 are, in all instances,

---

Insert Tables 1 and 2 about here

---

based on the narrative passage, while items 16 through 30 correspond to the expository passage. Omission of factor loadings less than .2 provides a clearer representation of the differences in loadings for items based on the two passage types. At all three grade levels, the expository items load primarily on the first factor while the narrative items load primarily on the second factor. Items that have slightly higher loadings on the "opposite" factor tend to be of a similar item type. For instance, on the Grade 3 test, items 7 and 12 are vocabulary items and

load to some extent on the "expository" dimension while items 19 and 22 are explicit items and load slightly on the "narrative" dimension. Although the items based on the two passages tend to primarily load on only one factor, the correlation between factors is quite high on each of the tests, suggesting that the dimensions assessed by items based on the two passage types are very similar.

DIMTEST analysis.

DIMTEST is a nonparametric statistical procedure based on Stout's (1987) theory of essential unidimensionality, the notion that unidimensional tests consist of items measuring one dominant dimension. This procedure tests the hypothesis that a particular set of test items is dimensionally distinct from another set of items (perhaps the remainder of the test). Depending on the purpose of the practitioner, DIMTEST can be used either to generally assess the lack of unidimensionality in a test or to confirm the existence of separate dimensions hypothesized a priori (Roussos, Stout & Marden, 1993). Because the objective of this investigation was to determine whether questions based on the narrative passages were dimensionally distinct from those based on the expository passages, the latter orientation was the one applicable here.

In the DIMTEST manual, Stout, Douglas, Junker & Roussos (1993) describe a single run of DIMTEST according to four steps, summarized as follows:

Step 1. Select a set of items, or assessment subtest, believed to be dimensionally distinct from another set of items on the test. Denote this subset of items AT1.

Step 2. From the set of items believed to be dimensionally distinct from AT1, DIMTEST selects a set of items which are similar in difficulty to those in AT1. This subset is the same size as AT1 and will be denoted AT2. The remaining set of items from which AT2 were drawn will serve as a partitioning subtest and is denoted PT.

Step 3. Examinees are partitioned into subgroups accoridng to their scores on the PT subtest.

Step 4. For each of the subgroups of examinees created in Step 3, compute two variance estimates for each of the two assessment subtests. The first variance estimate is the observed score variance for examinees of the given PT subgroup, while the second variance estimate is the estimated variance assuming unidimensionality. If, after aggregating over PT subgroups, the difference between the two variance estimates is greater for the AT1 subtest than the AT2 subtest, this is evidence of multidimensionality. The DIMTEST test statistic T is computed from these differences and has been proven to be asymptotically normally distributed with mean 0 and variance 1, assuming unidimensionality (Stout, 1987).

In the present analysis, items from either the narrative or expository portion of the test comprised AT1 whereas items based on the alternate passage were used for AT2 and PT.Because DIMTEST works with dichotomous item responses, each of the five yes/no responses to a question was treated as an item, resulting in a total of 75 items for each of the two passages. Due to the equal number of

narrative and expository questions on each IGAP reading test, only subsets of the total set of items from one passage could be used for AT1, so as to ensure a sufficient number of items for the PT subtest. Results from 4 DIMTEST runs at each grade level are presented in Table 3. At all grades, two runs used items from

---

Insert Table 3 about here

---

the narrative passage for AT1 and two runs used items from the expository passage for AT1. At grade 8, items based on the expository passage were consistently more difficult than those based on the narrative passage making it impossible to find 25 items of nearly equal difficulty for assignment to the AT1 and AT2 subtests. This is needed to control for bias and is discussed by Stout (1993). As a result, only 12 items were used for AT1. Results at all grade levels consistently suggest that items based on the narrative passages are dimensionally distinct from those based on the expository passages. It should be noted, however, that although the obtained p-values are quite small, this reflects our level of certainty that the two dimensions are distinct and not the extent to which the dimensions are correlated.

In addition to containing items based on different types of passages, the IGAP also contains items requiring different cognitive abilities. These can be classified into the categories explicit, inferential (lower-level and higher-level),

application transfer, and vocabulary. Although items of each type are included in each passage, the number of items of each type occasionally varies across the two passages. Additional DIMTEST runs were performed including items of only a particular type in part to determine if the results found above may have simply been an artifact of different types of questions being asked for the two passage types. Results are presented in Table 4. The results are also quite consistent and

---

Insert Table 4 about here

---

suggest that the dimensional distinction between items based on narrative vs. expository passages can be attributed to more than differences in item type. The one exception is with Grade 8 explicit items, although this can perhaps be attributed in large part to the small size of AT1.

NOHARM analysis.

When items are believed to measure more than one ability, it is frequently appropriate to consider a multidimensional model, such as the multidimensional normal ogive or multidimensional logistic models. The NOHARM program employs an approach introduced by McDonald (1985) in which "nonlinear harmonic" approximations to the normal ogive are obtained through least squares estimation of item parameters (Fraser & McDonald, 1988). The multidimensional

normal ogive model can be written as:

$$P\{y_j = 1 \mid \underline{\theta}\} = c_j + (1-c_j)N[d + \underline{a}_j'\underline{\theta}]$$

NOHARM uses the product-moment correlation matrix of items to produce estimates of both the difficulty (d) and discrimination (a_j) parameters and a residual matrix to examine goodness of fit. It can be used in fitting either restricted or non-restricted models and performs varimax and promax rotations of the factor loadings (Fraser & McDonald, 1988).

As in the DIMTEST analysis, the investigation here makes use of the dichotomous item responses, so that the test is considered to consist of 75 narrative and 75 expository items. Because of the response dependency noted by Wang (1994) and the strict assumption of local independence in item response theory, the items selected for analysis were those for which response dependency was low, namely those from clusters in which few examinees were found to have zero or three previous "yes" responses.

In addition to estimating item parameters, NOHARM can be used to estimate correlations between dimensions believed to underly performance on given sets of items. Fixing the narrative items along one dimension and the expository items along another dimension allows estimation of the correlation between the dimensions when the correlation parameter is left free to be estimated. Table 5 contains the results from several runs of NOHARM in which this correlation was

estimated. At all three grade levels, the correlation between the "narrative" and

---

Insert Table 5 about here

"expository" dimensions is quite high (.81 - .84) when items of all types are considered together. Also included in the table are the results of NOHARM runs in which only items of a particular type were included. Once again, the results suggest that the dimensions defined by the narrative and expository items are highly correlated. A mild exception occurs among the application-transfer type items, although this may be largely a function of the low number of items of this type on each of the tests.

Item parameter estimates obtained from NOHARM can be used to represent items graphically. Reckase has introduced a representation in which the characteristics of an item (e.g. discrimination and difficulty) can be used to represent items as vectors in a 2D space (Reckase, 1985). The direction of each vector is obtained from the discrimination parameter estimates and corresponds to the direction that is being best measured in the (θ_1,θ_2) space, while the length indicates the amount of discrimination provided by the item in that direction. Item difficulty is represented by the location of the tail of the vector from the origin. More difficult items are entirely inside the first quadrant while vectors representing easier items are found in the third quadrant. Figure 1 illustrates this

representation through a series of graphs. A 2-dimensional item response surface

---

Insert Figure 1 about here

---

can be used to represent performance on an item when the probability of correct response is influenced by two abilities. The item vector used to represent an item has direction equal to the direction of maximum slope, or gradient, of the surface. Item response surfaces can also be represented by equal-probability contours, in which case the item vector representing the item is drawn perpendicular to the parallel contours.

Figures 2, 3, and 4 display item vector plots for each of the three grade levels. If the narrative and expository items represented two separate dimensions, we would expect the two types of vectors to lie in different sectors of the latent space. It is evident from each of the plots that even when conditioning on item-type, there is substantial variability among items based just on the narrative or expository passages. The distinction between items based on the two passage types perhaps appears greatest at the grade 3 level, although even there the average composite directions appear quite similar.

Insert Figures 2, 3 and 4 about here

Conclusion

Both the prinicpal components factor analysis and DIMTEST results suggest that the narrative and expository passages are measuring dimensionally distinct skills. Whether this distinction is entirely responsible for the disparate equating results obtained by Hsu (1994) is not clear. Correlation estimates resulting from the promax rotation and NOHARM estimates suggest that the distinction between dimensions is not very large, indicating that separate scale scores based on the two passages would not likely provide much additional information. However, such results could be due to what many have claimed is a general lack of exposure to appropriate expository reading material in early school years. As schools increasingly adopt policies to improve this situation, it may be that the abilities become more clearly differentiated, lending support to the option of reporting separate scale scores.

References

Armbruster, B.B, Anderson, T.H. & Ostertag,G. (1989). Teaching text structure to improve reading and writing. The Reading Teacher, 43, 130-137.

Beck, I. L., & McKeown,M.G. (1989). Expository text for young readers: The issue of coherence. In W. Kintsch and L. B. Resnick (Eds.), Knowing, learning and instruction: Essays in honor of Robert Glaser. Hillsdale, NJ : Lawrence Erlbaum Assoc.

Englert, C.S. & Hiebert, E.H. (1989). Children's developing awareness of text structure in expository materials. Journal of Educational Psychology, 76, 65-74.

Fraser, C. & McDonald, R.P. (1988). NOHARM: Least squares item factor analysis. Multivariate Behavioral Research, 23, 267-269.

Hsu, Y. & Ackerman, T. (1994). Equating reading test scores that combine narrative and expository test formats. Paper present at the 1994 AERA Annual Meeting, New Orleans.

Illinois State Board of Education (1991). The Illinois Reading Assessment: Classroom connections. Springfield, IL.: Illinois State Board of Education.

Leslie, L. & Caldwell, J. (1988). The Qualitative Reading Inventory: Issues in the development of a reading diagnostic test. In S. McCormick & J. Zutell (Eds.), Cognitive and social perspectives for literacy research and

instruction. Thirty-Eighth Yearbook of the National Reading Conference.

Chicago: National Reading Conference Inc.

Piccolo, J. (1987). Expository test structure: Teaching and learning strategies. The Reading Teacher, 40, 838-847.

Roussos, L.A., Stout, W.F., & Marden, J.I. (1993). Analysis of the multidimensional structure of standardized tests using DIMTEST with hierarchical cluster analysis. Submitted for publication.

Spiro, R.A. & Taylor, B.M. (1980). On investigating children's transition from narrative to expository discourse: The multidimensional nature of psychological text classification. (Tech. Rep. No. 195). Urbana -Champaign, IL: University of Illinois, Center for the Study of Reading.

Stout, W.F. (1987). A nonparametric approach for assessing latent trait dimensionality. Psychometrika, 52, 589-617.

Stout, W.F. (1993). Overdispersion: A strikingly simple non-parametric basis for rigorous modeling and statistical assessment (via DIMTEST) of latent dimensionality. Paper presented at the 9th Annual University of Twente Invitational IRT Workshop.

Stout, W.F., Douglas, J., Junker, B., & Roussos, L.A. (1993). DIMTEST Manual. Unpublished manuscript available from W.F. Stout, Univerisity of Urbana-Champaign, Illinois.

Wang, C. & Ackerman, T. (1994). An examination of response dependency when there is more than one correct answer. Paper presented the the 1994 Annual AERA Meeting, New Orleans.

Figure Captions

Figure 1. Geometric Represention of Items as Vectors

Figure 2. Item Vector Plots for Grade 3 Items.

Figure 3. Item Vector Plots for Grade 6 Items.

Figure 4. Item Vector Plots for Grade 8 Items.

Table 1. Eigenvalues

Grade 3		Grade 6		Grade 8	
Factor	Eigenvalue	Factor	Eigenvalue	Factor	Eigenvalue
1	18.7650	1	14.7591	1	14.4361
2	3.0710	2	2.6121	2	2.9549
3	2.4388	3	2.2155	3	2.2257
4	1.9682	4	1.7403	4	1.2840
5	1.4621	5	1.0546	5	.9706
6	1.3370	6	.9341	6	.9319
7	1.1739	7	.8607	7	.8673
8	1.0819	8	.8295	8	.8198
9	1.0355	9	.7719	9	.7539
10	.9789	10	.7465	10	.7209
11	.8878	11	.7077	11	.6785
12	.8060	12	.6945	12	.6572

Rotation Method: Promax

Grade 3			Grade 6			Grade 8		
Item	Factor 1	Factor2	Item	Factor1	Factor2	Item	Factor1	Factor2
1	.	.51975	1	.	.27957	1	.	.
2	.	.40963	2	.21032	.	2	.	.40440
3	.	.41336	3	.	.40471	3	.	.30099
4	.	.44871	4	.25833	.23550	4	.	.39775
5	.	.47664	5	.21601	.27298	5	.	.32167
6	.	.54674	6	.34905	.	6	.	.37489
7	.	.40327	7	.	.58655	7	.	.43965
8	.	.44203	8	.	.29223	8	.	.58107
9	.	.53993	9	.21103	.24967	9	.	.53162
10	.	.44546	10	.	.56785	10	.	.34094
11	.	.40855	11	.	.47780	11	.	.53401
12	.26610	.36823	12	.	.55543	12	.	.54550
13	.	.40885	13	.	.61125	13	.	.
14	.	.33063	14	.	.55097	14	.	.38002
15	.	.34256	15	.21922	.29663	15	.	.40144
16	.38844	.	16	.28009	.	16	.54623	.
17	.44525	.	17	.56271	.	17	.49973	.
18	.40514	.	18	.51455	.	18	.	.
19	.35329	.26749	19	.38602	.	19	.46286	.
20	.42466	.	20	.47427	.	20	.64343	.
21	.52771	.	21	.49030	.	21	.53522	.
22	.23139	.36755	22	.55028	.	22	.47623	.
23	.51433	.	23	.44264	.	23	.45320	.
24	.65416	.	24	.46910	.	24	.51895	.
25	.59383	.	25	.54739	.	25	.45528	.
26	.56105	.	26	.50294	.	26	.69274	.
27	.63385	.	27	.44945	.	27	.46762	.
28	.57079	.	28	.46616	.	28	.30970	.
29	.56177	.	29	.30785	.	29	.40309	.
30	.59419	.	30	.38167	.	30	.43060	.

Inter-factor correlation: .68722

Inter-factor correlation: .67289

Inter-factor correlation: .64969

Table 3. DIMTEST Confirmatory Analysis for Narrative vs. Expository Items

Grade 3

AT1	PT & AT2	p-value
1,4,7,10,13,16,19,22,25, 28,31,34,37,40,43,46,49, 52,55,58,61,64,67,70,73	All Expository Items	.001158
1,2,5,6,7,9,13,18,22,27, 28,30,32,33,38,44,46, 47,48,51,55,60,64,66,72	All Expository Items	.000001
3,6,9,12,15,18,21,24,27, 30,33,36,39,42,45,48,51, 54,57,60,63,66,69,72,75	All Narrative Items	$<5 \times 10^{-7}$
2,6,9,12,13,17,22,24,26, 27,33,35,36,41,44,49,52, 53,57,59,60,65,66,72,74	All Narrative Items	$<5 \times 10^{-7}$

Grade 6

AT1	PT & AT2	p-value
1,4,7,10,13,16,19,22,25, 28,31,34,37,40,43,46,49, 52,55,58,61,64,67,70,73	All Expository Items	.001718
1,2,5,6,7,9,13,18,22,27, 28,30,32,33,38,44,46,47, 48,51,55,60,64,66,72	All Expository Items	.003330
3,6,9,12,15,18,21,24,27 30,33,36,39,42,45,48,51, 54,57,60,63,66,69,72,75	All Narrative Items	$<5 \times 10^{-7}$
2,6,9,12,13,17,22,24,26, 27,33,35,36,41,44,49,52, 53,57,59,60,65,66,72,74	All Narrative Items	.027472

Grade 8

AT1	PT & AT2	p-value
2,8,14,20,26,32,38, 44,50,56,62,68	All Expository Items	.001230
4,11,15,20,29,31,36, 44,48,59,68,70	All Expository Items	.000125
1,7,13,19,28,30,37,43, 49,55,61,67,73	All Narrative Items	$<5 \times 10^{-7}$
3,13,16,24,30,32,46, 50,57,66,68,75	All Narrative Items	$<5 \times 10^{-7}$

Table 4. DIMTEST Confirmatory Analysis for Narrative vs. Expository Items of a Particular Item Type

Grade 3

Item Type	AT1	AT2 & PT	p-value
Explicit	76,79,92,95,98,106, 109,112,115,123	34 Narrative Items (Explicit)	$<5 \times 10^{-7}$
Inferential	81,82,83,84,85,101, 102,103,104,105	31 Narrative Items (Inferential)	.000049
Application Transfer	31,32,33,34,35	25 Expository Items (App. Transfer)	.000064

Grade 6

Item Type	AT1	AT2 & PT	p-value
Explicit	11,16,19,21,23,27,31, 34,46,48,50,52,55	40 Expository Items (Explicit)	$<5 \times 10^{-7}$
Inferential (Type I)	76,78,80,97,99,136, 138,140	24 Narrative Items (Inferential-I)	$<5 \times 10^{-7}$

Grade 8

Item Type	AT1	AT2 & PT	p-value
Explicit	80,84,88,99,105,134	18 Narrative Items (Explicit)	.124447
Inferential	91,94,97,106,109,112, 115,118,121,124,127, 130	47 Narrative Items (Inferential)	$<5 \times 10^{-7}$
Application Transfer	146,148,150	10 Narrative Items (App. Transfer)	$<5 \times 10^{-7}$

Table 5. NOHARM Correlation Estimates Between Narrative and Expository Dimensions

All items

Grade	Correlation
3	.812
6	.840
8	.819

Explicit Items

Grade	Correlation
3	.752
6	.835
8	.803

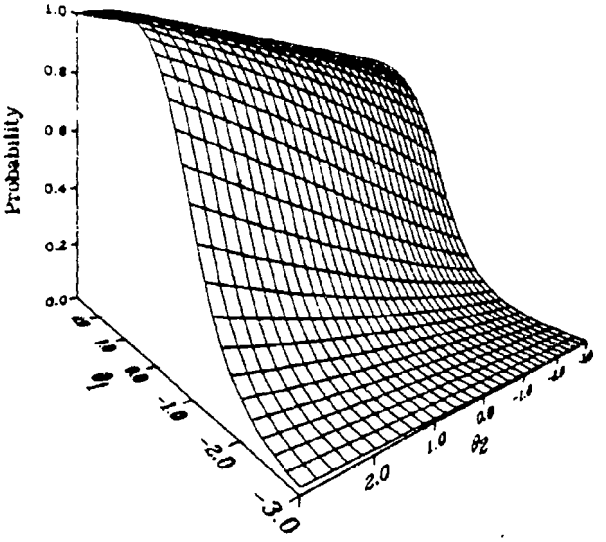
Inferential Items

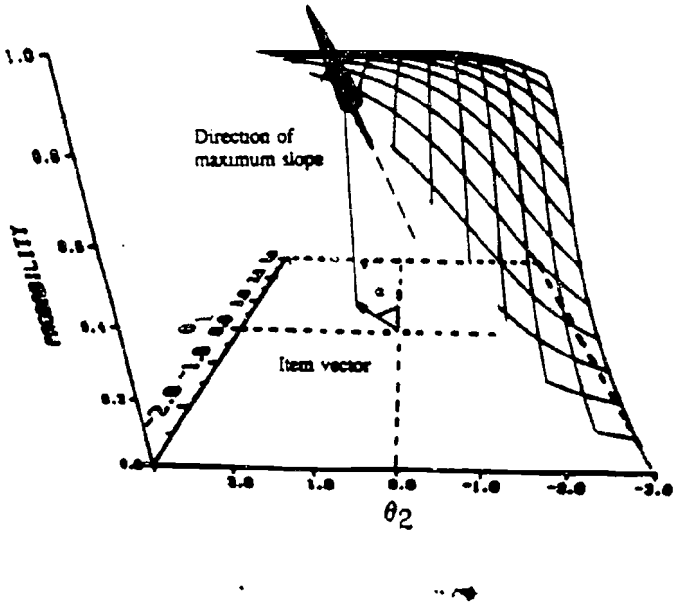
Grade	Correlation
3	.782
6	.800
8	.768

Application Transfer Items

Grade	Correlation
3	.678
6	.358
8	.648

Figure 1


Probability
1.0
0.8
0.6
0.4
0.2
0.0
-3.0
-2.0
-1.0
0.0
1.0
2.0
θ2


Direction of
maximum slope
Item vector
θ2

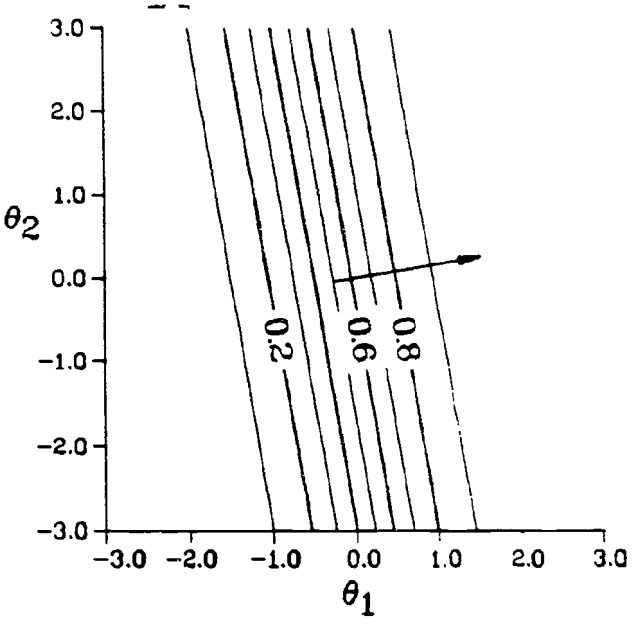

3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
θ2
0.2
0.6
0.8
−3.0 −2.0 −1.0 0.0 1.0 2.0 3.0
θ1

Figure 2

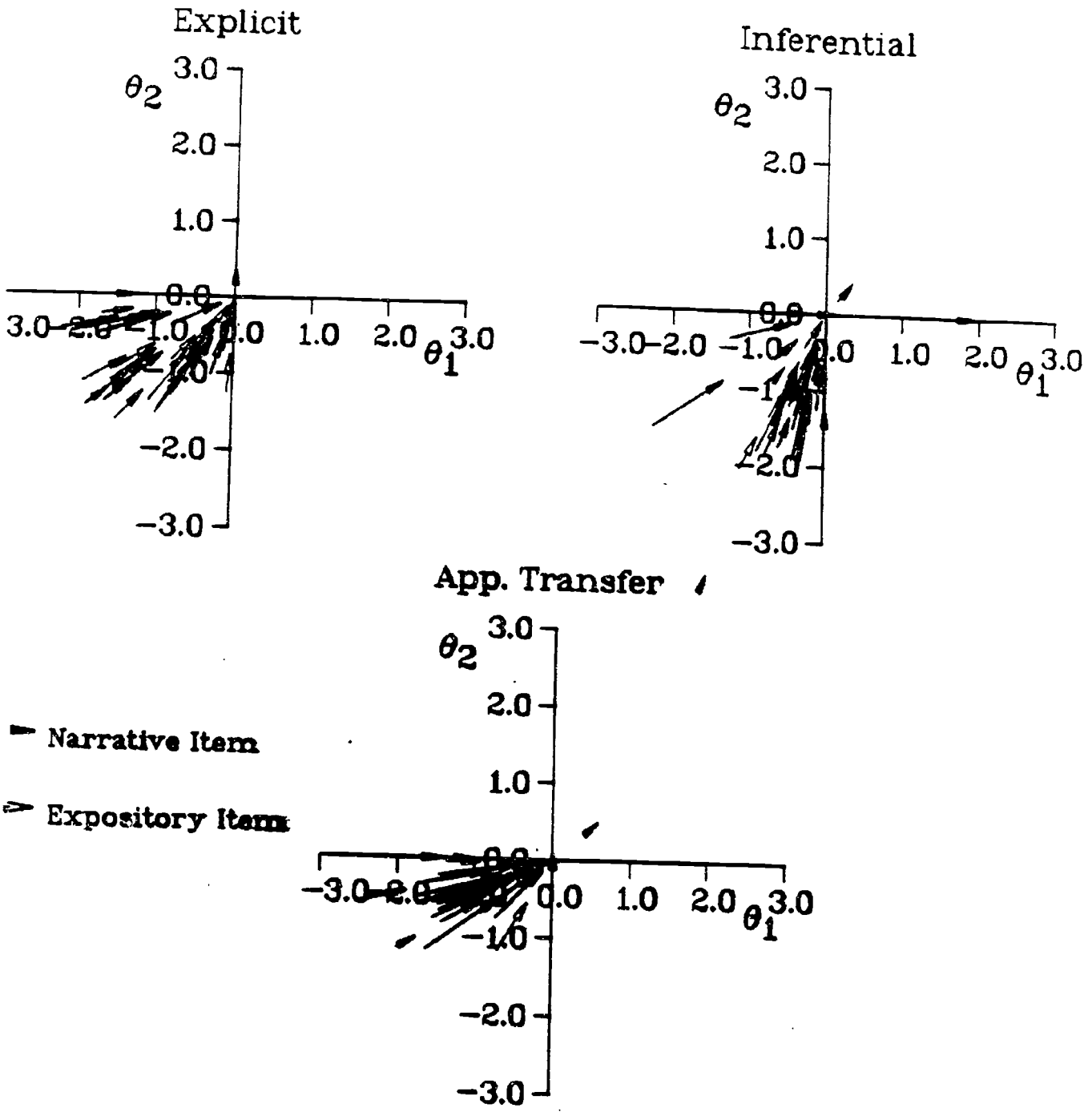

Explicit
Inferential
App. Transfer
θ_2
θ_1
3.0
2.0
1.0
0.0
−1.0
−2.0
−3.0
Narrative Item
Expository Item

Figure 3

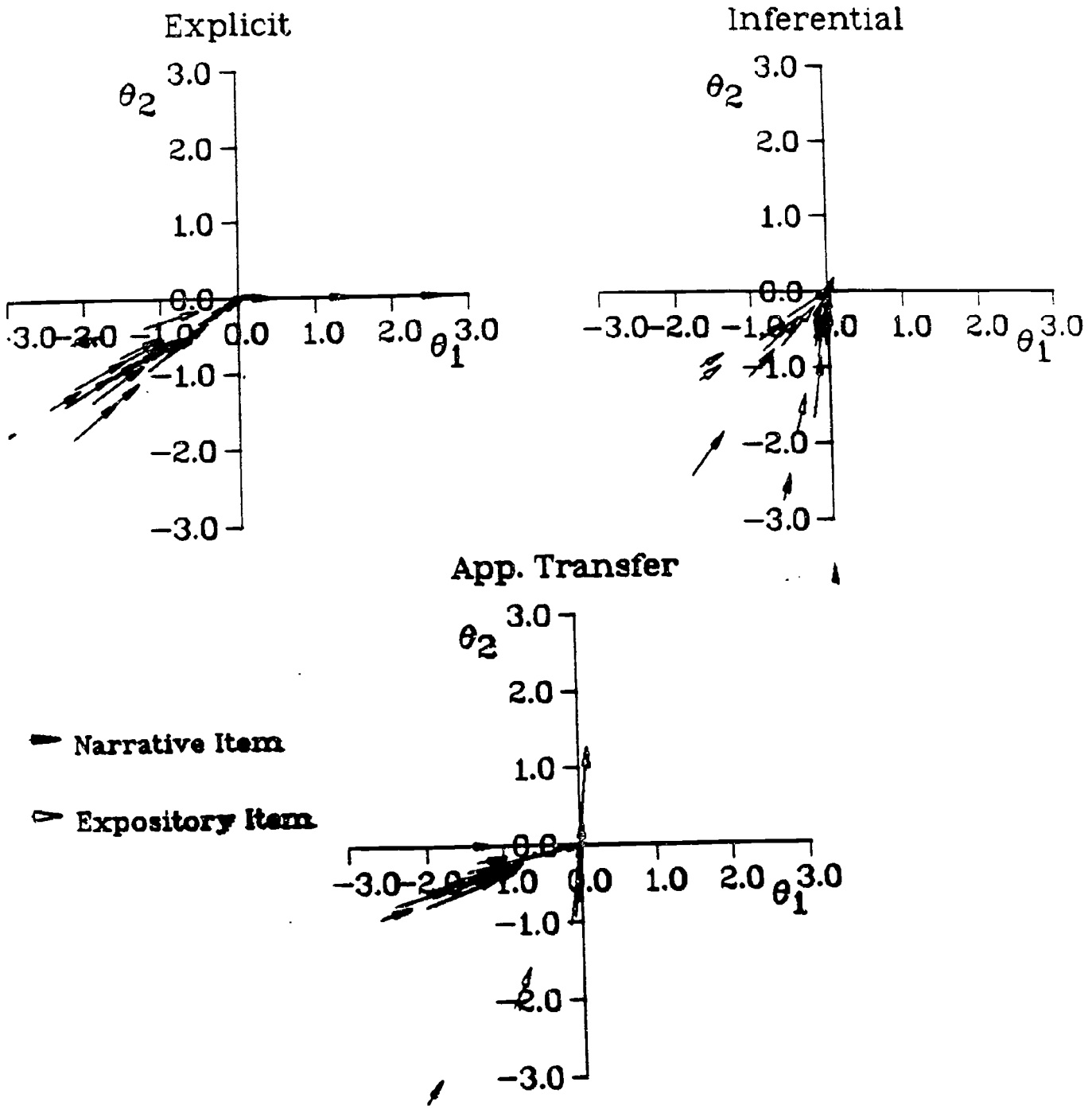
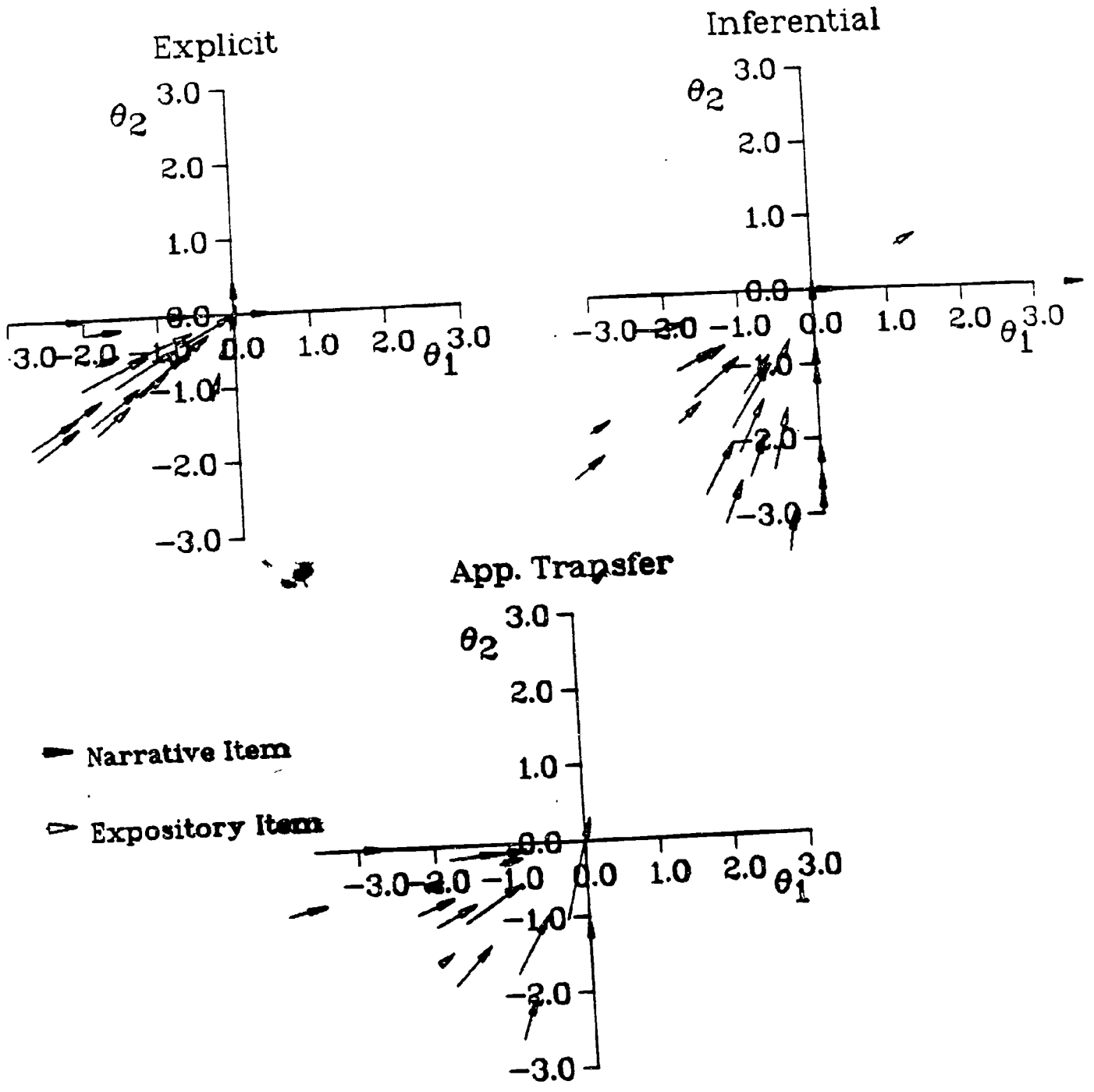

Explicit
Inferential
App. Transfer
θ_2
θ_1
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
Narrative Item
Expository Item

Figure 4


Explicit
Inferential
App. Transfer
θ_2
θ_1
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
Narrative Item
Expository Item