DOCUMENT RESUME

ED 299 321 TM 012 349

AUTHOR	Marzano, Robert J.; Jesse, Daniel M.
TITLE	A Study of General Cognitive Operations in Two Achievement Test Batteries and Their Relationship to Item Difficulty.
INSTITUTION	Mid-Continent Regional Educational Lab., Aurora, CO.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	Dec 87
CONTRACT	400-86-0002
NOTE	47p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Achievement Tests; Cognitive Measurement; *Cognitive Processes; Comparative Analysis; Construct Validity; *Difficulty Level; Elementary Secondary Education; Psychometrics; *Test Items; *Test Validity
IDENTIFIERS	Comprehensive Tests of Basic Skills; *General Cognitive Operations; Stanford Achievement Tests

ABSTRACT

In this study 6,942 items from two standardized achievement test batteries--the Stanford (Early School Achievement Battery, Achievement Battery, and Test of Academic Skills) and the Comprehensive Test of Basic Skills--were analyzed. Focus was on determining: (1) the extent to which these test batteries included general cognitive operations (GCOs) that are considered important for information processing; and (2) the relationship of the GCOs they contain to item difficulty. It was found that only nine of 22 GCOs were represented in the two tests--retrieval, reference, comparison and contrast, summarizing, inference, ordering, visual matching, transposing, and representing. When item difficulty was regressed on these nine GCOs, it was found that very little of the variance in item difficulty could be accounted for. Three possible interpretations for this weak relationship between the nine GCOs and item difficulty are discussed. Two interpretations imply that GCOs are a valid construct and are important to domain specific tasks. The third interpretation implies that GCOs are an invalid construct since cognition cannot be separated from content. A list of 39 references is attached. Four tables present test results and intercorrelations. (Author/SLD)

A STUDY OF GENERAL COGNITIVE OPERATIONS IN TWO ACHIEVEMENT TEST BATTERIES AND THEIR RELATIONSHIP TO ITEM DIFFICULTY

by

Robert J. Marzano

and

Daniel M. Jesse

Mid-continent Regional Educational Laboratory

Aurora, Colorado

December, 1987

This publication is based on work sponsored wholly, or in part, by the Office of Educational Research and Improvement, Department of Education, under Contract Number 400-86-0002. The content of this publication does not necessarily reflect the views of OERI or any other agency of the U.S. Government.

Abstract

In this study 6,942 items from two standardized achievement test batteries, the Stanford and the Comprehensive Test of Basic Skills, were analyzed to determine: 1) the extent to which they included general cognitive operations considered important for information processing, and 2) the relationship of the general cognitive operations they contain to item difficulty. It was found that only nine of 22 general cognitive operations were represented in the two tests. When item difficulty was regressed on the nine general cognitive operations it was found that very little of the variance in item difficulty could be accounted for. Three possible interpretation for this weak relationship between the nine general cognitive operations and item difficulty are discussed. Two interpretations imply that general cognitive operations are a valid construct and are important to domain specific tasks. The third implies that general cognitive operations are an invalid construct since cognition can not be separated from content.

Standardized testing is grounded on the assumption that underlying the ability to correctly answer test items are important cognitive structures and operations useful in more general contexts. The very process of insuring face validity by using a two-way specification table is an attempt to identify some of those cognitive structures and operations (Anastasi, 1982; Gronlund, 1977).

The most systematic attempts to determine the nature of test items have been statistical in nature rather than cognitively based. For example, much of the work in item response theory (IRT) can be characterized as an attempt to account for (statistically) but not label the underlying cognitive operations necessary to answer specific item types (Hambleton and Swaminathan, 1985; Trabin and Weiss, 1983). So too is the multi-dimensional item difficulty (MID) procedure (Reckase, 1985).

There have been, however, some attempts to analyze test items from a strict cognitive perspective. Among these were Adey and Harlen's (1986) Piagetian based analysis of test items specifically designed to measure the science process skills of 11-year-olds. They found that the developmental level of the cognitive tasks in such specially designed test items was a reliable predictor of item difficulty. Similarly O'Brien (1986) calibrated mathematics test items on the basis of the complexity of the cognitive processes necessary to complete the items. Tanner (1986) developed a college level test and found that the level of abstraction and level of cognitive processing involved in answering items accounted for a significant amount of variance in subjects' test scores. Similarly, using think-aloud protocols and a multi-method approach for examining the cognitive level of multiple choice items written by medical professors, Simpson and Cohen (1985) found that more cognitively complex items were more difficult.

Although the research findings, thusfar, indicate a relationship between the cognitive complexity of specially designed test items and their difficulty, there has been little systematic study of items on standardized test batteries to determine the types of cognitive operations involved and their relationship to item difficulty. Perhaps the closest study to this end was that conducted by Drum, Calfee and Cook (1980). They analyzed surface structure features on reading comprehension test items from the California Achievement Test, the Comprehensive Test of Basic Skills and the Sequential Test of Educational Progress. They found that such surface level syntactic characteristics as word length, propositional density and syntactic density accounted for as much as three fourths of the variance in item difficulty. However, they came to few conclusions regarding the underlying cognitive operations represented by these surface level characteristics.

Given the lack of a cognitively based analysis of items on standardized tests, the present study attempted to identify those general cognitive operations which are embedded in items from the various sections of two commonly used achievement batteries and to determine the relations of those cognitive operations to item difficulty as measured by *p* value. More specifically, we attempted to identify that general "procedural" knowledge necessary to answer items on standardized tests.

Many theorists have stressed the importance of the declarative/procedural distinction as it relates to information processing (Anderson, 1982, 1983; Paris and Lindauer, 1982). Declarative knowledge is factual in nature and includes such structures as concepts (Klausmeier, 1985), principles (Katz, 1976) and schemata (Rumelhart 1975, 1980). Procedural knowledge includes knowledge of processes and the conditions under which those processes should be used (Paris, Lipson and Wixson, 1983). It is procedural knowledge that we refer to as cognitive operations. General cognitive operations are those procedures which are not specific to any given domain. Rather they are used for academic tasks in more than one content area.

Recently there has been a great deal of interest in developing the general cognitive skills (general procedural knowledge) of students (Costa, 1985; Marzano, Brandt, Jones, Hughes, Rankin, Presseisen and Suhor, 1987; Paul, 1984). This interest is predicated on the assumption that a knowledge of and facility with such operations will help students in the processing of a wide variety of information. Some have argued that the curriculum of most content area courses and the standardized tests used to assess content area knowledge do not cover and consequently do not reinforce many of the cognitive operations important to the processing of information (Beyer, 1985). A related issue, then, to that of the relationship of general cognitive operations to item difficulty on standardized tests, is the extent to which standardized test items include many of the cognitive operations allegedly important to the processing of information. Consequently, in this study we attempted to answer two research questions: 1) to what extent do standardized test items include the use of general cognitive operations considered necessary for information processing, and 2) what is the relationship between the general cognitive operations found in standardized test items and item difficulty?

METHOD

Items from the following levels of the Stanford Early School Achievement Battery (197'), the Stanford Achievement Test (1973) and the Stanford Test of Academic Skills (1975) were analyzed for inclusion of general cognitive operations. All three tests were considered part of the overall Stanford battery:

<u>Stanford</u>

Stanford Early School Achievement Battery

K 1	Form E
K 2	Form E

Stanford Achievement Battery

P 1	Form E
P 2	Form E
P 3	Form E
I 1	Form E
I 2	Form E

Stanford Test of Academic Skills

A	Form E
Task 1	Form E
Task 2	Form E

Also analyzed for inclusion of the general cognitive operations were the following levels of the Comprehensive Test of Basic Skill (1984) henceforth referred to as the CTBS:

<u>CTBS</u>

Level A	Form U
Level B	Form U
Level C	Form U
Level D	Form U
Level E	Form U
Level F	Form U
Level G	Form U
Level H	Form U
Level J	Form U
Level K	Form U

Within each level selected subtests were analyzed. Spelling subtests were excluded from analyses on both batteries because it was assumed that given the similarity of format for all items (e.g., the examinee dictates a list of words of which students identify the correct spelling) there would be little or no variance in the type of cognitive operations used among items. The subtests analyzed on the Stanford and CTBS batteries were:

Stanford	CTBS
Language	Language
Listening	Listening
Mathematics	Mathematics
Reading Comprehension	Reading Comprehension
Science	Science
Social Science	Social Science
	Vocabulary
	Wordreading

General cognitive operations were operationally defined as those which can be used in more than one academic area. A list of general cognitive operations was developed by combining those identified by Costa, (1985), Marzano, et al (1987) and Ennis (1985). The combined list included 22 general cognitive operations. These are listed in Table 1.

Table 1 here.

To test the ability of raters to identify these cognitive operations in test items, a pilot study was conducted.

The Pilot Study

Each of the 22 general cognitive operations were operationally defined and sample items exemplifying each were constructed. Two raters were then trained to recognize these operations by constructing sample items for each of the 22 types. Once both raters could construct items which they agreed were illustrative of the 22 general cognitive operations, both raters independently analyzed all items on three sub-tests from a single level of the Stanford. The raters attempted to identify all operations in each item from the list of 22. Of the 22 only six were found on the three sub-tests analyzed. Inter-rater reliabilities were calculated for each of the six identified operations using Pearson Product Momement correlations. They ranged from .75 to .94 ($\underline{p} < .01$). The raters then jointly revised the operational definitions for the 22 cognitive operations. To test the effects of establishing these new criteria the raters again independently rated a single sub-test. Five of the 22 operations were found in that sub-test, and inter-rater reliabilities ranged from .84 to .98. These results were taken as strong evidence that general cognitive operations could be reliably identified, if present, on items from the Stanford and the CTBS.

Analysis of Items

Upon completion of the pilot study one rater analyzed all items from the selected subtests of the Stanford and CTBS. In all, 6,942 items were analyzed (3,775 on the Stanford and 3,167 on the CTBS) and nine general cognitive operations found within those items. They were: 1) retrieval, 2) reference, 3) comparison and contrast, 4) summarizing, 5) inference, 6) ordering, 7) visual matching, 8) transposing, and 9) representing

Retrieval refers to "calling up" declarative information not stated in an item into working memory (Anderson, 1983). That is, an item was coded as involving the general cognitive operation of retrieval if it required students to recall information not literally stated in the item. For example, if an item mentioned the concept carnival and then asked for information about carnivals not explicitly stated in the item or its support materials, it was assumed that students had to retrieve the information about carnivals from long term memory.

Reference refers to identifying information either explicitly stated in an item or from long term memory as cued by syntax, pronouns, synonyms or subordinate or superordinate terms (Halliday & Hasan, 1976). For example, if an item mentioned the term carnival and then later referred to explicit or implicit information about carnivals using a pronoun (e.g. *it*) it was assumed that the cognitive operation, reference, was necessary to answer the item.

Comparison and contrast refers to the process of identifying similar and/or dissimilar attributes between or among analogous terms (Stahl, 1985). For example, an item which required students to discern similarities or differences between or among concepts, principles and other cognitive structures (e.g. schemata) was coded as including the process of comparison and contrast.

Summarizing refers to the process of combining information parsimoniously into a cohesive statement (Brown, Campione and Day, 1981). It involves such heuristics as: selecting what is important, disregarding what is not and combining the selected information in some parsimonious way. An item which required students to construct or recognize a summary statement was coded as utilizing the process of summarizing.

Inferring is the process of inducing or deducing unstated information (Halpern, 1984). For example, generating or recognizing characteristics of subordinate concepts within some superordinate category would involve deductive inference; generating or recognizing a principle for which examples have been provided would involve inductive inference. Inference differs from retrieval in that it involves making new connections and associations among information in long term memory as opposed to simply retrieving connections and associations already stored.

Ordering is the process of identifying attributes of things in relative or absolute terms and ranking or sequencing them according to those attributes (Marzano et. al. 1987). For example, if an item required students to identify the best or worst element among a set it was assumed that the item involved the cognitive operation of ordering.

Transposing is the process of translating information from one code to another. It is based on the fundamental assumption within semiotics that the meaning of any piece of declarative or procedural information can be encoded in one or more sign systems (Eco, 1976). For example, if an item required a student to translate words to numbers or vise versa it was assumed that the item required the cognitive operation of transposing.

Representing is the process of creating a graphic or pictographic mental or visual representation of information. It is based on the assertion by theorist such as Paivio (1971) that information is processed in two primary forms, linguistic and nonlinguistic. For example, if an item required students to create a mental representation or diagram of information it was assumed that the item involved the operation of representing.

Visual matching is the process of linking a picture or symbol with a linguistic label. This too is based on a duel-coding theory of information storage which asserts that all information has two primary forms of storage. Visual matching occurs when information is presented linguistically and nonlinguistically and individuals are asked to match the linguistic representation with a nonlinguistic visual symbol or picture of the information.

Analysis of Data

A series of analyses of variance using a general linear model approach (Finn, 1974) was conducted to determine the nature and strength of the relationship between the independent variables (the types of general cognitive operations) and p-value. Separate analyses were conducted for each battery, the Stanford and the CTBS, 1) by grade level, 2) by subtest, and 3) by battery (all items collapsed within a battery).

Descriptive statistics (e.g. means, standard deviations and correlations) were also calculated for each level of analyses described above.

RESULTS

The overall mean p-value for the Stanford items was .74 and for the CTBS, was .67. Table 2 displays the percentages of occurrence for each of the nine cognitive operations by battery.

TABLE 2 here

It is interesting to note that retrieval and comparison and contrast appeared in every item. That is, all items require students to retrieve specific declarative information not actually stated, and all items required students to compare and contrast information.

Because these two operations were found in all items they were dropped from any subsequent analyses. Table 3 reports the intercorrelations for both batteries among the seven independent variables that were left in the analyses.

TABLE 3 here

Since there was little colinearity among the independent variables, all were included in subsequent analyses.

Separate analyses of variance were conducted using the seven independent variables described above and p-value as the dependent variable for: 1) each level within battery, 2) each subtest within battery, and 3) for each battery. Out of the 34 analyses conducted the multiple R's ranged from .011 to .507. Table 4 displays the multiple R's by level and by subtest for each battery.

Table 4 Here

The analyses of variance by battery (all items collapsed within a battery) yielded multiple R's of .105 and .187 for the Stanford and CTBS respectively with squared multiple R's of .027 and .035. Further analyses which included terms representing all possible two and three way interactions were also conducted at the battery level. These analyses increased the multiple R's very slightly to .176 and .191 respectively. Consequently further analyses were not conducted by level or subtest with interaction terms included.

DISCUSSION

There were two basic findings in this study: 1) the test batteries analyzed covered a minority of the general cognitive operations identified in the literature on general cognitive skills, and 2) those general cognitive operations which were found accounted for very little variance in student performance.

The first of these findings is not surprising. The literature base from which the list of general cognitive operations was drawn is based on the assumption that formalized schooling does not explicitly deal with (either in its testing procedures or its instructional procedures) those general cognitive abilities which are useful across a wide range of academic and non-academic tasks. For example, Sternberg (1985) has noted that intelligence is comprised of a number of "components" many of which are not part of formal schooling. Similarly, Gardner (1983) theorizes that there are multiple types of intelligence only a few of which are commonly instructed and assessed. It is no wonder, then, that a list of cognitive operations drawn from a theory base which assumes deficiencies in current education assessment practices would not be well represented on a standardized test. The second finding, however, is rather surprising

Given the importance that most procedurally based models of human cognition place on general cognitive operations, one would expect that those general cognitive operations which were found on the test items analyzed would account for more variance than that found in this study. There are a number of possible explanations for this. One is that students have learned the general cognitive operations found on the Stanford and CTBS to a level of automaticity prior to the time they take the tests. Specifically, Anderson (1983) and Fitts (1964) assert that an individual progresses through at least three stages when learning a cognitive operation -- the last stage of which is the autonomous stage. At this stage the execution of the cognitive operation occurs automatically requiring very little of the capacity of working memory (LaBerge and Samuels, 1974). Following the theory of automaticity, one could conclude that the general cognitive operations found on the Stanford and CTBS are important to answering test items but because they have been learned by students to a level of automaticity, their presence or absence in the process of answering a test item accounts for little of the variance in item difficulty.

Another explanation is that the general cognitive operations found in the test items analyzed were fairly low level examples of those operations. That is, some development psychologists (Fischer, 1980; Case, 1985) assert that any general cognitive operation can be executed at a number of levels ranging from very simple to quite complex. Following this line of reasoning one could conclude that the general operations found on the Stanford and CTBS are such simple versions of those operation, that most students can easily perform them, and they are, consequently, not major factors in the successful completion of test items.

Both of the previous explanations would be consistent with the previously cited findings of Adey and Harlan (1986), O'Brien (1986), Tanner (1986) and Simpson and Cohen (1985). Specifically, the interpretation that general cognitive operations found on standardized tests are either known by students at the level of automaticity or are low level examples of the operations, would not preclude the possibility of specially constructed items the difficulties of which are more dependent on the difficulty of the underlying cognitive operations.

Another perspective and possible explanation for the results is offered by Resnick (1983, in press) and Glaser (1984). They contend that general cognitive operations are meaningless in a practical sense unless considered in conjunction with the information which is being processed. That is, one cannot separate general cognitive operations from the content being tested. From this perspective, one could conclude that the general cognitive operations identified in the test batteries analyzed interacted with the content to such an extent that they can not be considered independently of content relative to their difficulty. This would explain why there was virtually no meaningful relationship among the seven independent variables and item difficulty. Item difficulty is not a function of the presence or absence per se of general cognitive operations. Rather, it is the nature of the information on which the cognitive operations are used which contributes to the difficulty of an item and, hence, student performance on standardized tests. However, this interpretation does not appear consistent with the previously cited studies which indicate that items can be constructed to reflect the difficulty of the underlying general cognitive operations.

SUGGESTIONS FOR FURTHER RESEARCH

The present study suggests that further research should be conducted on the interaction of general cognitive operations and domain specific declarative knowledge and the hierarchic nature relative to difficulty of domain specific declarative knowledge within academic content. Specifically there is a need to determine: 1) whether those general cognitive operations identified in the literature on thinking skills can be executed at different levels of difficulty independent of the content on which they are operating, and 2) the nature and difficulty of the declarative information on standardized tests independent of the cognitive operations which operate on that information.

REFERENCES

Adey, P. & Harlen, W. (1986). A Piagetian analysis of process skill test items, Journal of Research in Science Teaching, 23, 707-726.

Anastasi, A. (1982). Psychological testing. New York: Macmillan.

Anderson, J. (1982). Acquisition of cognitive skills. Psychological Review, 89, 369-406.

Anderson, J. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Beyer, B. (1985). Practical strategies for the direct teaching of thinking, in A. Costa (Ed.) Developing minds: A resource book for teaching thinking. Alexandria, VA: Association for Supervision and Curriculum Development.

Brown, A.L., Campione, J.C. & Day, J. (1981). Learning to learn: On training students to learn from texts. Educational Researcher, 10, 14-24.

Case, R. (1985). Intellectual development: Birth to adulthood. New York: Academic Press.

Costa, A. (1985). Teacher behaviors that enable student thinking, in A. Costa, (Ed.), Developing minds: A resource book for teaching thinking. Alexandria, VA: Association for Supervision and Curriculum Development.

Drum, P., Calfee, R. & Cook, L. (1980). The effects of surface structure variables on performance in reading comprehension. Reading Research Quarterly, 16, 486-513.

Ennis, R.H. (1985). Goals for a critical thinking curriculum. in A. Costa (Ed.) Developing minds: A resource book for teaching thinking. Alexandria, VA: Association for Supervision and Curriculum Development.

Eco, U. (1976). A theory of semiotics. Bloomington, IN: Indiana University Press.

Finn, J. (1974). A general model for multivariate analysis. New York: Holt and Rinehart & Winston, Inc.

Fischer, K.W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. Psychological Review, 87, 477-531.

Fitts, P.M. (1964). Perceptual - motor skill learning, in A. W. Melton, (Ed.). Categories of human learning. New York: Academic Press.

Gardner, H. (1983). Frames of mind: The theory of multiple intelligence. New York: Basic Books.

Glaser, R. (1984). Education and thinking: The role of knowledge, American Psychologist, 39, 93-104.

Gronlund, N.E. (1977). Constructing achievement tests. Englewood Cliffs, N.J. Prentice-Hall.

Halliday, M. & Hasan, R. (1976). Cohesion in English. London: Longman.

Halpern, D.F. (1984). Thought and knowledge: An introduction to critical thinking. Hillsdale, NJ: Erlbaum.

Hambleton, R.K. & Swaminathan, H. (1985). Item response theory. Boston: Kluwer, Nijhoft Publishing.

Katz, S.E. (1976). The effect of each of four instructional treatments on the learning of principles by children. Madison, WI: Wisconsin Research and Development Center for Conition Learning. The University of Wisconsin.

Klausmeier, H.J. (1985). Educational psychology. New York: Harper and Row.

LaBerge, D. and Samuels, S.J. (1974). Toward a theory of automatic information processing in reading comprehension, Cognitive Psychology, 6, 293-323.

Marzano, R., Brandt, R., Jones, B.F., Hughes, C., Rankin, S., Presseisen, B. & Suhor, C. (1987). Dimension of thinking. Alexandria, VA: Association for Supervision and Curriculum Development.

O'Brien, M.L. (1986). Calibrating item difficulty as the basis of prescriptive test theory. Studies in Educational Evaluation, 12, (1), 5-29, 59-69.

Paivio A. (1971). Imagery and verbal processes. New York: Holt.

Paris, S.G. & Lindauer, B.K. (1982). The development of cognitive skills during childhood, in B.W. Wolman (Ed.), Handbook of Developmental Psychology. Englewood Cliffs, NJ: Prentice-Hall.

Paris, S.G., Lipson, M.Y., & Wixson, K.K. (1983). Becoming a strategic reader, Contemporary educational psychology, 8, 293-315.

Paul, R. (1984). Critical thinking: Fundamental to education for a free society, Educational Leadership, 42, 1, 4-14.

Reckase, M. (April, 1985). The difficulty of test items that measure more than one ability. Paper presented at the annual meeting of the American Educational Research Association. Chicago, IL.

Resnick, L.B. (1983). Toward a cognitive theory of instruction, in S. Paris, G. Olson, & H. Stevenson (Eds.), Learning and motivation in the classroom. Hillsdale, NJ: Erlbaum.

Resnick, L.B. (in press). Education and learning. Pittsburg, PA: Learning Research and Development Center, University of Pittsburg.

Rumelhart, D. (1975). Notes on a schema for stories, in D. Bobrow & A. Collins (Eds.), Representation and understanding: Studies in cognitive science. New York: Academic Press.

Rumelhart, D. (1980). Schemata, The building blocks of cognition, in R.J. Spiro, B.C. Bruce, & W.I. Brewer (Eds.), Theoretical issues in reading comprehension. Hillsdale, NJ: Erlbaum.

Simpson, P. & Cohen, E. (March, 1985). Problem solving questions for multiple choice tests: A method for analyzing cognitive demands of items. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Stahl, R.J. (1985). Cognitive information processes and processing within a uniprocess superstructure/microstructure framework: A practical information based model. Tucson, AZ: University of Arizona.

Sternberg, R.J. (1985). Beyond IQ: A triarchic theory of human intelligence. New York: Cambridge University Press.

Tanner, D.E. (April, 1986). Achievement as a function of abstractness and cognitive level. Paper presented at the annual meeting of the American Educational Research Association. San Francisco, CA.

Trabin, T.E. & Weiss, D.J. (1983). The person response curve: Fit of individuals to item response theory models, in D.J. Weiss (Ed.) New horizons in testing. New York: Academic Press.

TEST REFERENCES

Comprehensive Test of Basic Skills. (1984). Monterey, CA: McGraw-Hill.

Stanford Achievement Battery. (1973). R. Madden, E.F. Gardner, H.C. Rudman, B. Karlsen & J.C. Merwin. Cleveland, OH: The Psychological Corporation.

Stanford Early School Achievement Battery. (1971). R. Madden & E.F. Gardner. Cleveland, OH: The Psychological Corporation.

Stanford Test of Academic Skills. (1975). E. Gardner, R. Callis, J. Merwin & R. Madden. Cleveland, OH: The Psychological Corporation

TABLE 1

List of General Cognitive Operations

categorizing

comparing and contrasting

creating analogies

creating metaphors

dialectic thinking

encoding

establishing criteria

extrapolating

identifying errors

identifying patterns and relationships

inferring

ordering

predicting

reference

restructuring

retrieving

representing

summarizing

transposing

valuing

verifying

visual matching

TABLE 2

Percentage of Occurence of Nine General Cognitive Operations

	Stanford	CTBS
Retrieval	100% N=3775	100% N=3167
Reference	16.8% N=635	26.5% N=839
Comparison/Contrast	100% N=3775	100% N=3167
Summarizing	2.4% N=92	3.6% N=115
Inference	5.7% N=215	7.1% N=224
Ordering	6.4% N=241	4.7% N=150
Visual matching	11.1% N=420	6.4% N=204
Transposing	4.4% N=167	6.3% N=200
Representing	7.0% N=266	2.6% N=81

TABLE 3

Intercorrelations Among General Cognitive Operations

Stanford

	Ref	Sum	Trans	Vim	Inf	Ord
Sum	-.07*					
Trans	-.05	-.03				
Vim	-.07*	.02	-.08*			
Inf	-.03	.01	-.05	.00		
Ord	.04	.24**	-.06	-.05	-.02	
Rep	-.10**	-.04	-.06	-.09**	.06*	.13**

CTBS

	Ref	Sum	Trans	Vim	Inf	Ord
Sum	.01					
Trans	-.08*	-.05				
Vim	-.06*	-.05	-.07*			
Inf	.09**	-.01	-.07*	-.01		
Ord	.01	.08*	-.06	.01	-.03	
Rep	-.05	-.02	-.04	-.04	.06	.12

Ref=reference
Sum=summarizing
Trans=transposing
Vim=visual matching
Inf=inferrence
Ord=ordering
Rep=representing
(*=<.05; **=<.01 2 -tailed test)

TABLE 4

Multiple Rs by Level and Subtest

Stanford

Level	N	Error df	R	R^2	P
K-1	189	182	.097	.009	N.S.
K-2	239	236	.165	.027	N.S.
P 1	309	303	.265	.070	**
P 2	353	346	.211	.044	**
P 3	476	468	.118	.014	N.S.
I 1	527	519	.144	.021	N.S.
I 2	537	530	.081	.007	N.S.
A	487	479	.157	.025	*
Task 1	329	323	.286	.086	**
Task 2	329	322	.278	.077	**

Subtest	N	Error df	R	R^2	P
Language	319	316	.236	.056	**
Listening	308	303	.225	.051	**
Math	841	836	.150	.022	*
Reading Comp	450	444	.280	.078	**
Science	456	450	.288	.083	**
Soc. Science	324	318	.149	.022	N.S.

(* <.05, ** <.01)

TABLE 4 (CONT.)

CTBS

Level	N	Error df	R	R^2	P
A	81	78	.011	.000	N.S.
B	101	96	.373	.139	**
C	145	140	.314	.099	**
D	260	254	.275	.075	**
E	30.	297	.250	.063	**
F	360	353	.304	.093	**
G	780	773	.240	.057	**
H	380	373	.198	.039	**
J	379	372	.120	.014	*
K	380	373	.204	.042	**

Subtest	N	Error df	R	R^2	P
Language	696	690	.164	.027	**
Listening	30	27	.056	.003	N.S.
Math	716	712	.230	.053	**
Reading Comp.	350	348	.274	.075	**
Science	295	288	.283	.080	**
Soc. Science	295	288	.323	.104	**
Vocab.	383	381	.108	.012	**
Word Study	172	170	.507	.257	*
Word Reading	450	444	.280	.078	**

(* <.05, ** <.01)

A STUDY OF GENERAL COGNITIVE OPERATIONS IN TWO ACHIEVEMENT TEST BATTERIES AND THEIR RELATIONSHIP TO ITEM DIFFICULTY

by
Robert J. Marzano
and
Daniel M. Jesse
Mid-continent Regional Educational Laboratory
Aurora, Colorado

December, 1987

This publication is based on work sponsored wholly, or in part, by the Office of Educational Research and Improvement, Department of Education, under Contract Number 400-86-0002. The content of this publication does not necessarily reflect the views of OERI or any other agency of the U.S. Government.

Abstract

In this study 6,942 items from two standardized achievement test batteries, the Stanford and the Comprehensive Test of Basic Skills, were analyzed to determine: 1) the extent to which they included general cognitive operations considered important for information processing, and 2) the relationship of the general cognitive operations they contain to item difficulty. It was found that only nine of 22 general cognitive operations were represented in the two tests. When item difficulty was regressed on the nine general cognitive operations it was found that very little of the variance in item difficulty could be accounted for. Three possible interpretation for this weak relationship between the nine general cognitive operations and item difficulty are discussed. Two interpretations imply that general cognitive operations are a valid construct and are important to domain specific tasks. The third implies that general cognitive operations are an invalid construct since cognition can not be separated from content.

Standardized testing is grounded on the assumption that underlying the ability to correctly answer test items are important cognitive structures and operations useful in more general contexts. The very process of insuring face validity by using a two-way specification table is an attempt to identify some of those cognitive structures and operations (Anastasi, 1982; Gronlund, 1977).

The most systematic attempts to determine the nature of test items have been statistical in nature rather than cognitively based. For example, much of the work i tem response theory (IRT) can be characterized as an attempt to account for (statistically) but not label the underlying cognitive operations necessary to answer specific item types (Hambleton and Swaminathan, 1985; Trabin and Weiss, 1983). So too is the multi-dimensional item difficulty (MID) procedure (Reckase, 1985).

There have been, however, some attempts to analyze test items from a strict cognitive perspective. Among these were Adey and Harlen's (1986) Piagetian based analysis of test items specifically designed to measure the science process skills of 11-year-olds. They found that the developmental level of the cognitive tasks in such specially designed test items was a reliable predictor of item difficulty. Similarly O'Brien (1986) calibrated mathematics test items on the basis of the complexity of the cognitive processes necessary to complete the items. Tanner (1986) developed a college level test and found that the level of abstraction and level of cognitive processing involved in answering items accounted for a significant amount of variance in subjects' test scores. Similarly, using think-aloud protocols and a multi-method approach for examining the cognitive level of multiple choice items written by medical professors, Simpson and Cohen (1985) found that more cognitively complex items were more difficult.

Although the research findings, thusfar, indicate a relationship between the cognitive complexity of specially designed test items and their difficulty, there has been little systematic study of items on standardized test batteries to determine the types of cognitive operations involved and their relationship to item difficulty. Perhaps the closest study to this end was that conducted by Drum, Calfee and Cook (1980). They analyzed surface structure features on reading comprehension test items from the California Achievement Test, the Comprehensive Test of Basic Skills and the Sequential Test of Educational Progress. They found that such surface level syntactic characteristics as word length, propositional density and syntactic density accounted for as much as three fourths of the variance in item difficulty. However, they came to few conclusions regarding the underlying cognitive operations represented by these surface level characteristics.

Given the lack of a cognitively based analysis of items on standardized tests, the present study attempted to identify those general cognitive operations which are embedded in items from the various sections of two commonly used achievement batteries and to determine the relations of those cognitive operations to item difficulty as measured by p value. More specifically, we attempted to identify that general "procedural" knowledge necessary to answer items on standardized tests.

Many theorists have stressed the importance of the declarative/procedural distinction as it relates to information processing (Anderson, 1982, 1983; Paris and Lindauer, 1982). Declarative knowledge is factual in nature and includes such structures as concepts (Klausmeier, 1985), principles (Katz, 1976) and schemata (Rumelhart 1975, 1980). Procedural knowledge includes knowledge of processes and the conditions under which those processes should be used (Paris, Lipson and Wixson, 1983). It is procedural knowledge that we refer to as cognitive operations. General cognitive operations are those procedures which are not specific to any given domain. Rather they are used for academic tasks in more than one content area.

Recently there has been a great deal of interest in developing the general cognitive skills (general procedural knowledge) of students (Costa, 1985; Marzano, Brandt, Jones, Hughes, Rankin, Presseisen and Suhor, 1987; Paul, 1984). This interest is predicated on the assumption that a knowledge of and facility with such operations will help students in the processing of a wide variety of information. Some have argued that the curriculum of most content area courses and the standardized tests used to assess content area knowledge do not cover and consequently do not reinforce many of the cognitive operations important to the processing of information (Beyer, 1985). A related issue, then, to that of the relationship of general cognitive operations to item difficulty on standardized tests, is the extent to which standardized test items include many of the cognitive operations allegedly important to the processing of information. Consequently, in this study we attempted to answer two research questions: 1) to what extent do standardized test items include the use of general cognitive operations considered necessary for information processing, and 2) what is the relationship between the general cognitive operations found in standardized test items and item difficulty?

METHOD

Items from the following levels of the Stanford Early School Achievement Battery (1971), the Stanford Achievement Test (1973) and the Stanford Test of Academic Skills (1975) were analyzed for inclusion of general cognitive operations. All three tests were considered part of the overall Stanford battery:

<u>Stanford</u>

Stanford Early School Achievement Battery

K 1	Form E
K 2	Form E

Stanford Achievement Battery

P 1	Form E
P 2	Form E
P 3	Form E
I 1	Form E
I 2	Form E

Stanford Test of Academic Skills

A	Form E
Task 1	Form E
Task 2	Form E

Also analyzed for inclusion of the general cognitive operations were the following levels of the Comprehensive Test of Basic Skill (1984) henceforth referred to as the CTBS:

CTBS

Level A	Form U
Level B	Form U
Level C	Form U
Level D	Form U
Level E	Form U
Level F	Form U
Level G	Form U
Level H	Form U
Level J	Form U
Level K	Form U

Within each level selected subtests were analyzed. Spelling subtests were excluded from analyses on both batteries because it was assumed that given the similarity of format for all items (e.g., the examinee dictates a list of words of which students identify the correct spelling) there would be little or no variance in the type of cognitive operations used among items. The subtests analyzed on the Stanford and CTBS batteries were:

Stanford	CTBS
Language	Language
Listening	Listening
Mathematics	Mathematics
Reading Comprehension	Reading Comprehension
Science	Science
Social Science	Social Science
	Vocabulary
	Wordreading

General cognitive operations were operationally defined as those which can be used in more than one academic area. A list of general cognitive operations was developed by combining those identified by Costa, (1985), Marzano, et al (1987) and Ennis (1985). The combined list included 22 general cognitive operations. These are listed in Table 1.

Table 1 here.

To test the ability of raters to identify these cognitive operations in test items, a pilot study was conducted.

The Pilot Study

Each of the 22 general cognitive operations were operationally defined and sample items exemplifying each were constructed. Two raters were then trained to recognize these operations by constructing sample items for each of the 22 types. Once both raters could construct items which they agreed were illustrative of the 22 general cognitive operations, both raters independently analyzed all items on three sub-tests from a single level of the Stanford. The raters attempted to identify all operations in each item from the list of 22. Of the 22 only six were found on the three sub-tests analyzed. Inter-rater reliabilities were calculated for each of the six identified operations using Pearson Product Momement correlations. They ranged from .75 to .94 ($p < .01$). The raters then jointly revised the operational definitions for the 22 cognitive operations. To test the effects of establishing these new criteria the raters again independently rated a single sub-test. Five of the 22 operations were found in that sub-test, and inter-rater reliabilities ranged from .84 to .98. These results were taken as strong evidence that general cognitive operations could be reliably identified, if present, on items from the Stanford and the CTBS.

Analysis of Items

Upon completion of the pilot study one rater analyzed all items from the selected subtests of the Stanford and CTBS. In all, 6,942 items were analyzed (3,775 on the Stanford and 3,167 on the CTBS) and nine general cognitive operations found within those items. They were: 1) retrieval, 2) reference, 3) comparison and contrast, 4) summarizing, 5) inference, 6) ordering, 7) visual matching, 8) transposing, and 9) representing

Retrieval refers to "calling up" declarative information not stated in an item into working memory (Anderson, 1983). That is, an item was coded as involving the general cognitive operation of retrieval if it required students to recall information not literally stated in the item. For example, if an item mentioned the concept carnival and then asked for information about carnivals not explicitly stated in the item or its support materials, it was assumed that students had to retrieve the information about carnivals from long term memory.

Reference refers to identifying information either explicitly stated in an item or from long term memory as cued by syntax, pronouns, synonyms or subordinate or superordinate terms (Halliday & Hasan, 1976). For example, if an item mentioned the term carnival and then later referred to explicit or implicit information about carnivals using a pronoun (e.g. it) it was assumed that the cognitive operation, reference, was necessary to answer the item.

Comparison and contrast refers to the process of identifying similar and/or dissimilar attributes between or among analogous terms (Stahl, 1985). For example, an item which required students to discern similarities or differences between or among concepts, principles and other cognitive structures (e.g. schemata) was coded as including the process of comparison and contrast.

Summarizing refers to the process of combining information parsimoniously into a cohesive statement (Brown, Campione and Day, 1981). It involves such heuristics as: selecting what is important, disregarding what is not and combining the selected information in some parsimonious way. An item which required students to construct or recognize a summary statement was coded as utilizing the process of summarizing.

Inferring is the process of inducing or deducing unstated information (Halpern, 1984). For example, generating or recognizing characteristics of subordinate concepts within some superordinate category would involve deductive inference; generating or recognizing a principle for which examples have been provided would involve inductive inference. Inference differs from retrieval in that it involves making new connections and associations among information in long term memory as opposed to simply retrieving connections and associations already stored.

Ordering is the process of identifying attributes of things in relative or absolute terms and ranking or sequencing them according to those attributes (Marzano et. al. 1987). For example, if an item required students to identify the best or worst element among a set it was assumed that the item involved the cognitive operation of ordering.

Transposing is the process of translating information from one code to another. It is based on the fundamental assumption within semiotics that the meaning of any piece of declarative or procedural information can be encoded in one or more sign systems (Eco, 1976). For example, if an item required a student to translate words to numbers or vise versa it was assumed that the item required the cognitive operation of transposing.

Representing is the process of creating a graphic or pictographic mental or visual representation of information. It is based on the assertion by theorist such as Paivio (1971) that information is processed in two primary forms, linguistic and nonlinguistic. For example, if an item required students to create a mental representation or diagram of information it was assumed that the item involved the operation of representing.

Visual matching is the process of linking a picture or symbol with a linguistic label. This too is based on a duel-coding theory of information storage which asserts that all information has two primary forms of storage. Visual matching occurs when information is presented linguistically and nonlinguistically and individuals are asked to match the linguistic representation with a nonlinguistic visual symbol or picture of the information.

Analysis of Data

A series of analyses of variance using a general linear model approach (Finn, 1974) was conducted to determine the nature and strength of the relationship between the independent variables (the types of general cognitive operations) and p-value. Separate analyses were conducted for each battery, the Stanford and the CTBS, 1) by grade level, 2) by subtest, and 3) by battery (all items collapsed within a battery).

Descriptive statistics (e.g. means, standard deviations and correlations) were also calculated for each level of analyses described above.

RESULTS

The overall mean p-value for the Stanford items was .74 and for the CTBS, was .67. Table 2 displays the percentages of occurrence for each of the nine cognitive operations by battery.

TABLE 2 here

It is interesting to note that retrieval and comparison and contrast appeared in every item. That is, all items require students to retrieve specific declarative information not actually stated, and all items required students to compare and contrast information.

Because these two operations were found in all items they were dropped from any subsequent analyses. Table 3 reports the intercorrelations for both batteries among the seven independent variables that were left in the analyses.

TABLE 3 here

Since there was little colinearity among the independent variables, all were included in subsequent analyses.

Separate analyses of variance were conducted using the seven independent variables described above and p-value as the dependent variable for: 1) each level within battery, 2) each subtest within battery, and 3) for each battery. Out of the 34 analyses conducted the multiple R's ranged from .011 to .507. Table 4 displays the multiple R's by level and by subtest for each battery.

Table 4 Here

The analyses of variance by battery (all items collapsed within a battery) yielded multiple R's of .105 and .187 for the Stanford and CTBS respectively with squared multiple R's of .027 and .035. Further analyses which included terms representing all possible two and three way interactions were also conducted at the battery level. These analyses increased the multiple R's very slightly to .176 and .191 respectively. Consequently further analyses were not conducted by level or subtest with interaction terms included.

DISCUSSION

There were two basic findings in this study: 1) the test batteries analyzed covered a minority of the general cognitive operations identified in the literature on general cognitive skills, and 2) those general cognitive operations which were found accounted for very little variance in student performance.

The first of these findings is not surprising. The literature base from which the list of general cognitive operations was drawn is based on the assumption that formalized schooling does not explicitly deal with (either in its testing procedures or its instructional procedures) those general cognitive abilities which are useful across a wide range of academic and non-academic tasks. For example, Sternberg (1985) has noted that intelligence is comprised of a number of "components" many of which are not part of formal schooling. Similarly, Gardner (1983) theorizes that there are multiple types of intelligence only a few of which are commonly instructed and assessed. It is no wonder, then, that a list of cognitive operations drawn from a theory base which assumes deficiencies in

Both of the previous explanations would be consistent with the previously cited findings of Adey and Harlan (1986), O'Brien (1986), Tanner (1986) and Simpson and Cohen (1985). Specifically, the interpretation that general cognitive operations found on standardized tests are either known by students at the level of automaticity or are low level examples of the operations, would not preclude the possibility of specially constructed items the difficulties of which are more dependent on the difficulty of the underlying cognitive operations.

Another perspective and possible explanation for the results is offered by Resnick (1983, in press) and Glaser (1984). They contend that general cognitive operations are meaningless in a practical sense unless considered in conjunction with the information which is being processed. That is, one cannot separate general cognitive operations from the content being tested. From this perspective, one could conclude that the general cognitive operations identified in the test batteries analyzed interacted with the content to such an extent that they can not be considered independently of content relative to their difficulty. This would explain why there was virtually no meaningful relationship among the seven independent variables and item difficulty. Item difficulty is not a function of the presence or absence per se of general cognitive operations. Rather, it is the nature of the information on which the cognitive operations are used which contributes to the difficulty of an item and, hence, student performance on standardized tests. However, this interpretation does not appear consistent with the previously cited studies which indicate that items can be constructed to reflect the difficulty of the underlying general cognitive operations.

SUGGESTIONS FOR FURTHER RESEARCH

The present study suggests that further research should be conducted on the interaction of general cognitive operations and domain specific declarative knowledge and the hierarchic nature relative to difficulty of domain specific declarative knowledge within academic content. Specifically there is a need to determine: 1) whether those general cognitive operations identified in the literature on thinking skills can be executed at different levels of difficulty independent of the content on which they are operating, and 2) the nature and difficulty of the declarative information on standardized tests independent of the cognitive operations which operate on that information.

REFERENCES

Adey, P. & Harlen, W. (1986). A Piagetian analysis of process skill test items, Journal of Research in Science Teaching, 23, 707-726.

Anastasi, A. (1982). Psychological testing. New York: Macmillan.

Anderson, J. (1982). Acquisition of cognitive skills. Psychological Review, 89, 369-406.

Anderson, J. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Beyer, B. (1985). Practical strategies for the direct teaching of thinking, in A. Costa (Ed.) Developing minds: A resource book for teaching thinking. Alexandria, VA: Association for Supervision and Curriculum Development.

Brown, A.L., Campione, J.C. & Day, J. (1981). Learning to learn: On training students to learn from texts. Educational Researcher, 10, 14-24.

Case, R. (1985). Intellectual development: Birth to adulthood. New York: Academic Press.

Costa, A. (1985). Teacher behaviors that enable student thinking, in A. Costa, (Ed.), Developing minds: A resource book for teaching thinking. Alexandria, VA: Association for Supervision and Curriculum Development.

Drum, P., Calfee, R. & Cook, L. (1980). The effects of surface structure variables on performance in reading comprehension. Reading Research Quarterly, 16, 486-513.

Ennis, R.H. (1985). Goals for a critical thinking curriculum. In A. Costa (Ed.) Developing minds: A resource book for teaching thinking. Alexandria, VA: Association for Supervision and Curriculum Development.

Eco, U. (1976). A theory of semiotics. Bloomington, IN: Indiana University Press.

Finn, J. (1974). A general model for multivariate analysis. New York: Holt and Rinehart & Winston, Inc.

Fischer, K.W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. Psychological Review, 87, 477-531.

Fitts, P.M. (1964). Perceptual - motor skill learning, in A. W. Melton, (Ed.). Categories of human learning. New York: Academic Press.

Gardner, H. (1983). Frames of mind: The theory of multiple intelligence. New York: Basic Books.

Glaser, R. (1984). Education and thinking: The role of knowledge, American Psychologist, 39, 93-104.

Gronlund, N.E. (1977). Constructing achievement tests. Englewood Cliffs, N.J. Prentice-Hall.

Halliday, M. & Hasan, R. (1976). Cohesion in English. London: Longman.

Halpern, D.F. (1984). Thought and knowledge: An introduction to critical thinking. Hillsdale, NJ: Erlbaum.

Hambleton, R.K. & Swaminathan, H. (1985). Item response theory. Boston: Kluwer, Nijhoft Publishing.

Katz, S.E. (1976). The effect of each of four instructional treatments on the learning of principles by children. Madison, WI: Wisconsin Research and Development Center for Conition Learning. The University of Wisconsin.

Klausmeier, H.J. (1985). Educational psychology. New York: Harper and Row.

LaBerge, D. and Samuels, S.J. (1974). Toward a theory of automatic information processing in reading comprehension, Cognitive Psychology, 6, 293-323.

Marzano, R., Brandt, R., Jones, B.F., Hughes, C., Rankin, S., Presseisen, B. & Suhor, C. (1987). Dimension of thinking. Alexandria, VA: Association for Supervision and Curriculum Development.

O'Brien, M.L. (1986). Calibrating item difficulty as the basis of prescriptive test theory. Studies in Educational Evaluation, 12, (1), 5-29, 59-69.

Paivio A. (1971). Imagery and verbal processes. New York: Holt.

Paris, S.G. & Lindauer, B.K. (1982). The development of cognitive skills during childhood, in B.W. Wolman (Ed.), Handbook of Developmental Psychology. Englewood Cliffs, NJ: Prentice-Hall.

Paris, S.G., Lipson, M.Y., & Wixson, K.K. (1983). Becoming a strategic reader, Contemporary educational psychology, 8, 293-316.

Paul, R. (1984). Critical thinking: Fundamental to education for a free society, Educational Leadership, 42, 1, 4-14.

Reckase, M. (April, 1985). The difficulty of test items that measure more than one ability. Paper presented at the annual meeting of the American Educational Research Association. Chicago, IL.

Resnick, L.B. (1983). Toward a cognitive theory of instruction, in S. Paris, G. Olson, & H. Stevenson (Eds.), Learning and motivation in the classroom. Hillsdale, NJ: Erlbaum.

Resnick, L.B. (in press). Education and learning. Pittsburg, PA: Learning Research and Development Center, University of Pittsburg.

Rumelhart, D. (1975). Notes on a schema for stories, in D. Bobrow & A. Collins (Eds.), Representation and understanding: Studies in cognitive science. New York: Academic Press.

Rumelhart, D. (1980). Schemata, The building blocks of cognition, in R.J. Spiro, B.C. Bruce, & W.I. Brewer (Eds.), Theoretical issues in reading comprehension. Hillsdale, NJ: Erlbaum.

Simpson, P. & Cohen, E. (March, 1985). Problem solving questions for multiple choice tests: A method for analyzing cognitive demands of items. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Stahl, R.J. (1985). Cognitive information processes and processing within a uniprocess superstructure/microstructure framework: A practical information based model. Tucson, AZ: University of Arizona.

Sternberg, R.J. (1985). Beyond IQ: A triarchic theory of human intelligence. New York: Cambridge University Press.

Tanner, D.E. (April, 1986). Achievement as a function of abstractness and cognitive level. Paper presented at the annual meeting of the American Educational Research Association. San Francisco, CA.

Trabin, T.E. & Weiss, D.J. (1983). The person response curve: Fit of individuals to item response theory models, in D.J. Weiss (Ed.) New horizons in testing. New York: Academic Press.

TEST REFERENCES

Comprehensive Test of Basic Skills. (1984). Monterey, CA: McGraw-Hill.

Stanford Achievement Battery. (1973). R. Madden, E.F. Gardner, H.C. Rudman, B. Karlsen & J.C. Merwin. Cleveland, OH: The Psychological Corporation.

Stanford Early School Achievement Battery. (1971). R. Madden & E.F. Gardner. Cleveland, OH: The Psychological Corporation.

Stanford Test of Academic Skills. (1975). E. Gardner, R. Callis, J. Merwin & R. Madden. Cleveland, OH: The Psychological Corporation

TABLE 1

List of General Cognitive Operations

categorizing
comparing and contrasting
creating analogies
creating metaphors
dialectic thinking
encoding
establishing criteria
extrapolating
identifying errors
identifying patterns and relationships
inferring
ordering
predicting
reference
restructuring
retrieving
representing
summarizing
transposing
valuing
verifying
visual matching

TABLE 2

Percentage of Occurence of Nine General Cognitive Operations

	Stanford	CTBS
Retrieval	100% N=3775	100% N=3167
Reference	16.8% N=635	26.5% N=839
Comparison/Contrast	100% N=3775	100% N=3167
Summarizing	2.4% N=92	3.6% N=115
Inference	5.7% N=215	7.1% N=224
Ordering	6.4% N=241	4.7% N=150
Visual matching	11.1% N=420	6.4% N=204
Transposing	4.4% N=167	6.3% N=200
Representing	7.0% N=266	2.6% N=81

TABLE 3

Intercorrelations Among General Cognitive Operations

Stanford

	Ref	Sum	Trans	Vim	Inf	Ord
Sum	-.07*					
Trans	-.05	-.03				
Vim	-.07*	.02	-.08*			
Inf	-.03	.01	-.05	.00		
Ord	.04	.24**	-.06	-.05	-.02	
Rep	-.10**	-.04	-.06	-.09**	.06*	13**

CTBS

	Ref	Sum	Trans	Vim	Inf	Ord
Sum	.01					
Trans	-.08*	-.05				
Vim	-.06*	-.05	-.07*			
Inf	.09**	-.01	-.07*	-.01		
Ord	.01	.08*	-.06	.01	-.03	
Rep	-.05	-02	-.04	-.04	.06	12

Ref=reference
Sum=summarizing
Trans=transposing
Vim=visual matching
Inf=inferrence
Ord=ordering
Rep=representing
(*=<.05; **=<.01 2 -tailed test)

TABLE 4

Multiple Rs by Level and Subtest

Stanford

Level	N	Error df	R	R^2	P
K-1	189	182	.097	.009	N.S.
K-2	239	236	.165	.027	N.S.
P 1	309	303	.265	.070	**
P 2	353	346	.211	.044	**
P 3	476	468	.118	.014	N.S.
I 1	527	519	.144	.021	N.S.
I 2	537	530	.081	.007	N.S.
A	487	479	.157	.025	*
Task 1	329	323	.286	.086	**
Task 2	329	322	.278	.077	**

Subtest	N	Error df	R	R^2	P
Language	319	316	.236	.056	**
Listening	308	303	.225	.051	**
Math	841	836	.150	.022	*
Reading Comp	450	444	.280	.078	**
Science	456	450	.288	.083	**
Soc. Science	324	318	.149	.022	N.S.

(* <.05, ** <.01)