COMPARING DIFFICULTY SEQUENCE OF KEYMATH-REVISED TEST ITEMS FOR THE NORMING SAMPLE AND FOR STUDENTS REFERRED FOR LEARNING DISABILITY ASSESSMENT

BY

DAN OWEN HOOKER

Bachelor of Science

Oklahoma State University

Stillwater, Oklahoma

1980

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
May, 1990

COMPARING DIFFICULTY SEQUENCE OF KEYMATH-REVISED TEST ITEMS FOR THE NORMING SAMPLE AND FOR STUDENTS REFERRED FOR LEARNING DISABILITY ASSESSMENT

Thesis Approved:

Paul G Warden

Thesis Advisor

Joseph Pearl

J Barbara Wilkinson

Norman N. Durham

Dean of the Graduate College

ACKNOWLEDGMENTS

I wish to express my gratitude to all those people who have helped me with this thesis, and throughout my graduate program. Special thanks go to my major advisor Dr. Paul Warden for his guidance, suggestions, and friendship. Thanks also go to my other committee members; Barbara Wilkinson, and Joseph Pearl, for the influence they had upon me personally and during coursework which made me want them as committee members. The Help of the staff at American Guidance Service is sincerely appreciated.

My parents, Edward and Marjory Hooker, and my siblings, Leigh, Ann, and Scott, have been a constant source of friendship, encouragement, and support. Thanks go to them for everything they do. Special thanks also go to Mark Savoca, Mark Lynch, Vanessa Tigert, Dale Fromberger, Bill Key, Joanne Puryear, and Ross Atkinson, who provided inspirational examples of academic pursuit for me to emulate. My sincere thanks go to Jane Harris who's growth promoting insights and friendship have been invaluable.

Special thanks also go to those who have shared in the unique challenges that have presented themselves concurrent to my thesis work; particularly the few who have shown unwavering "Sucking of marrow" not only on Crazy Alice, but in life as well.

Friends like Bruce Perry, Michael Simpson, Robert Careless, and Gary Gabrel have been greatly appreciated for demonstrating with their lives that there is rich life before, during, and after thesis. Finally, thanks go to the many people, whose numbers are too great to mention individually, who have assisted me throughout my thesis and coursework.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

The Keymath Revised, (Connally, 1988) is an individually administered diagnostic instrument, designed to inventory essential mathematics skills. It was published in 1988 by the American Guidance Service Inc. This instrument is designed for use across a broad age range, from kindergarten through grade nine. It is available in two alternate forms. Both forms contain 258 items presented in free-standing test books. The examiner shows the item prompt and asks the associated question. The subject responds, and the examiner marks the item correct or incorrect in the separate test record.

The Keymath-R (Connally, 1988) is based on thirteen content areas. It is designed to identify hierarchies of concepts and skills in each of the thirteen content areas. Each of these content areas is assessed by a corresponding subtest. Each of the subtests is comprised of three or four domains. Each domain is represented by six test items which represent a subgroup of skills within the subtest. Each of these domains is said in the test manual to be of nearly equal instructional importance. Domains are designed to facilitate assessment of math skills below the subtest level. These thirteen subtests are broken down into three areas and are presented below with the domains that contribute to each.

I. Basic Concepts

1. Numeration
    - Numbers 0-9
    - Numbers 0-99
    - Numbers 0-999
    - Multi-digit numbers
2. Rational Numbers
    - Fractions
    - Decimals
    - Percents
3. Geometry
    - Spatial/attribute relations
    - Two-dimensional shapes
    - Coordinates/transformations
    - Three-dimensional shapes

II. Operations

4. Addition
    - Models and basic facts
    - Algorithms: whole numbers
    - Adding rational numbers
5. Subtraction
    - Models and basic facts
    - Algorithms: whole numbers
    - Subtracting rational numbers
6. Multiplication
    - Models and basic facts
    - Algorithms: whole numbers
    - Multiplying rational numbers

7. Division

    Models and basic facts
    Algorithms: whole numbers
    Dividing rational numbers

8. Mental Computation

    Computation chains
    Whole numbers
    Rational numbers

III. <u>Applications</u>

9. Measurement

    Comparisons
    Non-standard units
    Standard units: length, area
    Standard units: weight, capacity

10. Time and Money

    Identifying passage of time
    Using clocks and clock units
    Monetary amounts to $1
    Monetary amounts to $100

11. Estimation

    Whole and rational numbers
    Measurement
    Computation

12. Interpreting Data

    Charts and tables
    Graphs
    Probability and statistics

13. Problem Solving

    Solving routine problems
    Understanding non-routine problems
    Solving non-routine problems

The Keymath-R (Connally, 1988) yields four levels of diagnostic scores. Each successive level offers increasingly specific information about the testee's mathematics ability. The first level is the Total Test Score which combines scores from all of the subtests and yields a standard score, grade and age equivalent, percentile rank, stanine, and normal curve equivalent, (NCE). The second level is the Area Score, again yielding standard score, grade and age equivalent, percentile rank, stanine, and NCE. These scores correspond to the testee's performance in the three areas listed above. The third level is the Subtest Score. This level yields scaled scores for each subtest, (mean of 10, standard deviation of 3), and percentile ranks. The fourth level is the Domain, which yields a raw score.

In order to reduce the number of items administered to each subject, the Keymath-R (Connally, 1988) is designed such that all 258 items need not be administered. The first subtest is started at a suggested item number appropriate for the subject's grade level. A basal of three consecutive correct responses is established and items are administered until a ceiling of three consecutive incorrect responses is reached. The starting item on subsequent subtests is based on where a basal was established on the first subtest. It is assumed that the subject would respond correctly to all items below the basal, and respond incorrectly to all items above the ceiling.

Statement of the Problem

Use of a basal and ceiling system is intended to help avoid fatigue by reducing the number of items administered. However; obtaining accurate scores is dependent upon administering all items that are appropriate for the testee. To insure that all appropriate items are administered, the items must be accurately sequenced according to difficulty. Inaccurate sequencing can result in basals that are too high or in ceilings that are too low. Inaccurate sequencing can also result in basals that are too low, or ceilings that are too high, requiring administration of excess items.

Establishing a basal requires three correct responses in a row. In theory, if the items are correctly sequenced, and the starting point is correctly estimated, each subject will correctly answer the first three items administered. If any one of the first three items is incorrectly answered, because it is too difficult and should have been administered later in the sequence, then the examiner must work backwards in the sequence until three items are correctly answered. If the third item administered is too difficult, and incorrectly answered, the examiner must seek to establish the basal by administering the item preceding the first item administered. If the second administered item is improperly sequenced and incorrectly answered, then the examiner must seek to establish a basal by administering the two items which precede the first item administered. If the first item administered is incorrectly answered, the examiner must administer the three preceding items in attempt to establish a basal. Any one item alone which, based on its relative difficulty, is placed too early in the sequence of items, and then incorrectly responded to, can result in the need for administration of three additional items, unnecessarily increasing administration time.

An item which based on its relative difficulty is placed too far along in the sequence of administration may result in a basal that is inaccurately high. If the third item administered is correctly answered because it is incorrectly sequenced, and if a correctly sequenced item would have been incorrectly answered, the result would be the establishing of a basal that is too high, and the subject would receive credit for a correct response which would not have received credit had the items been correctly sequenced.

Establishing a ceiling requires that three items in a row be answered incorrectly. When this occurs, no further items would be administered from that subtest. If one of the three ceiling items, relative to its difficulty, is placed too far along in the sequence, and if it is correctly answered, when a correctly sequenced item would have been incorrectly answered, this would result in the need to administer three further items. If this occurs on the first of the three ceiling items, obtaining a ceiling would require administering one further item in the sequence which otherwise would not have been administered. If this occurs on the third ceiling item, three additional items must be administered in order to establish a ceiling. In this case the subject would have been exposed to three items which would not have been administered had the items been correctly sequenced. An item which was administered too high in the sequence and correctly answered, would result in a subject being credited with one correct answer which may have been incorrectly responded to had the difficulty sequencing been correct. Administration of further items which the subject would not have been exposed to would allow for the possibility of additional correct responses, which would further inflate the subject's score. If additionally administered items are responded to incorrectly, then the negative result is limited to the administration of excess items which lengthens the overall time of administration and increases the subject's fatigue.

In theory, when a subject incorrectly responds to an item, if there is correct difficulty sequencing, the subject will likely respond incorrectly to the next two items and a ceiling will be established. If based on its relative difficulty, an item is placed too low in the administration sequence, it may be responded to incorrectly. If this item is the first of three consecutive incorrect responses, an early ceiling will have been achieved. Additional items which could have elicited correct responses would not be administered, resulting in an inaccurately low score.

The Keymath-R manual (Connally, 1988) provides a table with the difficulty sequences of all items in each of the thirteen subtests. These difficulty sequences were determined using standardization data. However, the test booklet was sequenced for administration before standardization. The sequence was based upon intuitive assumptions, without the benefit of the standardization data. This item order of administration which was not sequenced empirically may be resulting in inaccurate basals and ceilings.

A student's score on the Keymath- R (Connally, 1988) is often used to help make educational placement decisions. A Keymath-R score can be used to compare with an expected score, based on a student's score on an IQ test. A discrepancy between these two scores indicates that the student is not achieving at his ability level, and if the discrepancy is large enough, it is often seen as evidence of a learning disability. If test items are not correctly sequenced according to their relative difficulties, scores may not accurately reflect a student's mathematical ability, and discrepancy between the expected score and the achievement score may not be accurate. Any placement decisions which utilized these scores might be based on inaccurate test data. The revised Keymath is a new instrument. It's acceptance rests in part upon the wide spread acceptance of the original Keymath. The Keymath-R has not yet received much research attention. In order to gain insight into the accuracy of scores that it yields, the Keymath-R should be subjected to the scrutiny of statistical as well as practical analysis.

Clinical use of the Keymath suggests that when used with a population referred for learning disability, item difficulty sequence may be significantly different than that which is reported in the testing manual. To insure that accurate basals and ceilings are being established for referred populations, research should be conducted to verify the item difficulty sequencing.

Research Question

Will the difficulty sequencing of test items on the Keymath-R be different for a learning disabled population (hereafter referred to as the referred population) than for the norming sample.

CHAPTER II

REVIEW OF THE LITERATURE

Keymath

The original form of the Keymath (Connally, 1971) was published by the American Guidance Service Inc. The norming sample consisted of 1222 subjects, in grades K through 7, from 42 different schools, in eight different states. The sample of schools included a range of geographic and racial representation from rural and urban areas. Weighting was used to make the sample conform to U.S. proportions obtained on variables of community size and race.

Reliability coefficients for grades K through 7 were obtained from split-half analysis. Split-half measures were adjusted for length by the Spearman-Brown formula. Total test reliability coefficients range from .94 to .97 across all grade levels. Subtest reliability coefficients ranged from .23 to .90 across all grade levels. (Connally, 1971)

As early as 1973, the Keymath was receiving attention as a very useful diagnostic tool. Bannatyne (1973) commented that the Keymath was well thought out and nicely constructed. It was also normed on a sufficiently large sample and it has good reliability and validity. It is particularly useful because deficit areas are noted in considerable detail, facilitating precise remedial prescription writing. Bannatyne (1973) stated that the Keymath should become a standard part of the test battery of everyone concerned with evaluating and treating LD students.

The Keymath has been compared to the California Achievement Test, and found to offer noteworthy advantages. It was found to measure more of the current Math curriculum, (Tinney 1975) and also requires neither reading nor writing. (Tinney 1975, Kratochwill and Demuth, 1976) In comparison to other diagnostic instruments the Keymath is particularly well standardized, reliable, and valid. The data clearly supports the use of the Keymath as a diagnostic measure of math functioning among LD students. (Greenstein and Strain, 1977)

Connolly, Nachtman, and Prichett (cited in Kraochwill and Demuth, 1976) stated that by the mid '70s the Keymath was one of the most common math tests used by educators at both elementary and secondary levels. The literature is replete with praise and support for use of the Keymath when diagnosing learning disabilities. The Keymath could be used with confidence for learning disability screening, diagnosis, and research. (McCullough and Zaremba, 1979) Wide usage of the Keymath coupled with positive support in the literature has made the Keymath an important part of assessment of special learning problems.

Keymath-Revised

The Keymath-Revised (Connally, 1988) was published in 1988. It was reviewed in the Fall issue of the Council for Educational Diagnostic Services Newsletter (Nicholson, 1988). When compared to the original Keymath, most of the desirable original traits are retained and new features are included. Standardization is sound. Reliability and validity coefficients are at acceptable levels. The Keymath-R was found to be very useful, particularly for assessing learning disabled students (Nicholson, 1988).

Rena Lewis (1989), reviewed the Keymath-R and said that the purpose of achievement tests is to provide information about a child's academic achievement in relation to other children in the same age group or grade. Lewis (1989) stated that the Keymath-Revised does that quite well.

Lewis (1989) noted that it is unclear whether handicapped individuals were included in the Keymath-R standardization sample; however, she stated that it is likely that samples taken from regular classes will include mainstreamed handicapped individuals. Lewis (1989) noted that the norms on the Keymath-R are greatly improved over the original Keymath, giving norm referencing rather than grade referencing as the reason for improvement.

In 1989 the Keymath-Revised was reviewed in the Journal of Psychoeducational Assessment (Huebner, 1989). It was said to appear to be an excellent diagnostic instrument for measuring mathematics achievement. The manual was said to be excellent, technical characteristics are impressive, and the entire test was described as user friendly. (Huebner, 1989)

Measurement Principles.

Use of a ceiling and basal system requires that items be sequenced according to their difficulty level. Difficulty is defined as the proportion of students responding correctly to an item. The higher the proportion is, the easier the item is. Conversely, the lower the proportion is, the harder the item is. The range of difficulty levels is from .00 to 1.00. A .00 difficulty level indicates that none of the sample correctly responded to the item. A 1.00 difficulty level indicates that none of the sample incorrectly answered the item. (Sax, 1989)

CHAPTER III

METHOD

Subjects

A sample of students was selected from a small rural midwest school, grades 6-12. The students selected had been diagnosed as having a learning disability or had been referred for learning disability assessment. In order to obtain an N of 40, additional students were randomly selected from files of learning disabled students in a second small rural midwest school, grades 6-12.

Procedures

During the 1988-89 school year the Keymath-Revised was administered to each of these students. Data was collected from each of the forty protocols. Each of the 258 items on a selected protocol were recorded as having been answered correctly or incorrectly. Items below an individual's basal were recorded as having been answered correctly, and items above an individual's ceiling were recorded as having been answered incorrectly. This data was statistically analyzed to determine item difficulty. Test items within each subtest were then placed in sequence according to their difficulty. Each resulting subtest item difficulty sequence was compared to the sequence presented in the Keymath-Revised test manual (Connally, 1988).

Instruments

The Keymath-Revised differs from the original Keymath in several ways. There are thirteen subtests on the Keymath-R, and fourteen subtests on the Keymath. In the Keymath-R there is a Rational Numbers subtest which includes not just fractions as in the original Keymath, but also decimals and percents. The Time subtest and the Money subtest from the original Keymath were combined to form a single subtest in the Keymath-Revised. The Mental Computation subtest was expanded to include not only the original mental computation chains, but other mental computation problems as well. A new subtest was developed to measure estimation skills. A new subtest was developed to measure the ability to interpret data; and a new subtest was developed to measure problem solving ability. The Numerical Reasoning and Missing Elements subtests from the original Keymath were not included in the Keymath-Revised. The Keymath -Revised was expanded from the 209 items on the original Keymath to 258 items. The alternate forms which are available with the Keymath-R were not available with the original Keymath (Connally, 1988).

Perhaps the most important change in the Keymath has to do with the types of scores that it yields. The Original

Keymath yielded only grade equivalents scores, while the Keymath-Revised yields standard scores, and percentile ranks in addition to grade equivalent scores. (Lewis, 1989)

The Keymath-Revised 1988 edition (Connally, 1988) includes a manual in which Reliability and Validity Data is presented. A reliability coefficient of .80 or higher is generally considered acceptable. (Satler, 1982) Alternative-form reliability coefficients were computed from grade-based scaled scores for subtests, and from grade-based standard scores for the areas and the total test. Correlations between form A and form B range from .50s to .70s for the subtests. They fall in the low .80s for the areas and average .90 for the total test.

Split-half reliability was obtained for the Keymath-R subtests by correlating odd and even test items. The Spearman-Brown formula was used to obtain estimates for the full-length test. Split-half reliability coefficients for the subtests, across the K-9 grades, range mostly in the .70s to the .80s. Coefficients fall mostly in the mid to high .90s for the areas, and coefficients for the total test fall mostly in the mid to high .90s.

The Keymath-R test manual presents evidence of test validity from three categories, including developmental change, internal consistency, and correlation with other tests. It is expected that scores on the Keymath-R will increase as the grade level of the student increases. This is based on the expectation that students in each higher grade will have been exposed to and learned more mathematics than students in lower grades. With minor exceptions, mean performance levels on the Keymath-R have been found to increase with grade level.

It is expected that scores that contribute to a particular area score will correlate more closely than scores from a different area. The test manual presents data indicating subtest scores correlate most highly with their respective area scores.

The test manual also presents coefficients of correlations between the Keymath-R and other mathematics achievement tests. Correlation coefficients for Total Test scores on the Keymath-R and Total Test scores on the Keymath range from the .80s to the mid .90s. Total test correlation between the Keymath-R and the Comprehensive Tests of Basic Skills is .66. Total test correlation between the Iowa Tests of Basic Skills and the Keymath-R is .76. These correlations indicate that the Keymath is measuring content which is similar to that measured by other widely used mathematics achievement tests.

CHAPTER IV

RESULTS

The purpose of this study was to determine if the difficulty sequence of test items on the Keymath-Revised, for a sample referred for learning disability testing, would be different from the difficulty sequence for the norming sample. Table I (see appendix) taken from the Keymath-R manual, shows the difficulty sequence for the norming sample. Table II (see appendix) shows the item difficulty sequence obtained in this study for a sample of referred students. Items are listed with their corresponding difficulty proportions.

Analysis

The total number of subjects in the sample (N = 40) places considerable limits on the data collected. Item difficulty is the proportion of individuals who correctly responded to an item. An N of 40 would make all item difficulty proportions a certain percent of 40. Each individual who correctly answers an item increases the proportion by 0.025. Therefore the item difficulty can not be determined more precisely than 0.025 intervals. It would be desirable to have a much larger sample such that smaller differences in difficulty could be measured. The limitations can be seen in Table II (see Appendix) where as many as seven items from a single subtest showed identical difficulty proportions. It is doubtful that the difficulty is truly identical for each of these items, but the small sample does not allow for more precise measurement.

A comparison of Tables I and II indicate that the item difficulty sequence yielded by this study is not the same as that presented in the Keymath-R Manual (Connally, 1988). The difficulty sequence obtained in this study shows that across all 13 subtests there are 12 items that differ by three or more positions from that sequence presented in the test manual. (see Table III in appendix) Two items differ by four positions, one item differs by five positions, and one item differs by six positions.

As previously addressed, items which are not placed in sequence according to difficulty can effect the test results by changing ceilings and basals. The student is thereby given an opportunity to respond to additional items, improving his score when correctly responding. If administration of additional items does not improve a student's score, the negative effect is limited to extending administration time which increases examiner and student fatigue.

Supplementary Analysis

A supplementary analysis of the 12 items noted in Table III indicates that several of these items had a meaningful impact upon time of administration as well as the student's scores. This occurred primarily when according to their difficulty proportion, items were placed too high in the sequence, and a ceiling was not achieved due to a correct response to the item in question. Across the entire sample, items 3, 6, and 7, from the Geometry subtest, and items 10, 7, and 5, from the Addition, Subtraction, and Multiplication subtests, respectively, had little meaningful effect on the testing. On the Mental Computation subtest, the difficulty proportion corresponding to item 12, indicated that it was placed too far along in the sequence. In the study sample 7 students failed to reach a ceiling which would have been reached had the response to item 12 been incorrect. This was determined only if the two previous items had been incorrectly responded to, and then the response to item 12 had been correct, or if the items on either side of item 12 were incorrectly answered, and item 12 was correctly answered. Three of the seven students went on to correctly respond to further questions. Four of the seven did not correctly answer further questions, but they were required to attempt answering further items as a result of having not reached a ceiling which included item 12.

Data from the sample indicated that on the Time and Money subtest, item 1 was placed too high in the sequence, but this did not appear to have a meaningful effect on the testing. Sample data indicates that item 11 was placed too high in the sequence. Three of the 40 sample students would have reached a ceiling had they not correctly answered item 11. All three of the students went on to correctly answer further items thereby increasing their scores as well as administration time.

Analysis of sample data indicates that in the Interpreting Data subtest, item number 9 is placed too high in the sequence. There were 6 students out of 40 who failed to reach a ceiling as a result of correctly answering item 9. Five of these students went on to correctly answer further questions.

In the Problem Solving subtest, sample data indicates that item 7 is placed too low in the sequence. This did not appear to have a meaningful effect on the testing. Sample data indicates that item 8 was placed too high in the sequence. Items 7 and 8 together had a meaningful effect on the scores of three sample students, each going on to correctly respond to further items. Had item 8 been incorrectly responded to, these students would have reached a ceiling, and administration would have discontinued.

Among the 40 sample protocols, 17 were effected in a meaningful way by at least one item from table III. Twelve of these subjects went on to correctly respond to further items which they would not have been exposed to had they incorrectly responded to the item in question, thereby reaching a ceiling.

CHAPTER V

SUMMARY AND CONCLUSIONS

This study involved collecting Keymath-Revised test scores from a sample of referred students. A difficulty proportion was obtained for each of these items indicating the percent of the sample that correctly responded to the item. Items were then placed in sequence of difficulty. This difficulty sequence was then compared to the sequence yielded by the norming sample (presented in the Keymath-Revised test manual (Connally,1988). This comparison indicated that there was a difference in the two sequences, with 12 items differing by 3 or more positions in the difficulty sequence.

It is not possible to confidently attribute this difference solely to the sample being a referred population. This difference could be a result of the individual school math curriculum, which placed emphasis on different math skills. The small size of this study is small, and the difficulty sequence yielded could be a sample specific characteristic.

Supplementary Analysis

A close examination of the individual sample protocols indicated that several of the items listed in Table III appeared to have a meaningful impact not only on scores, but also on the length of test administration. Seventeen of the 40 sample protocols were effected in a meaningful way by one or more of the items in Table III. This effect was limited to extending administration time with 5 of the 17 students. However 12 of the 17 appear to have received higher scores as well as extended administration time.

Conclusions

This study compared difficulty sequences of Keymath-R items for the norming sample, to the sequence yielded by the study sample of referred students. Findings indicated that the difficulty sequence may be different for a referred population. This should be studied further, with a much larger sample.

One of the uses of the Keymath-R scores is for determining achievement/ability discrepancy, and using this as evidence of a learning disability. This study indicates that individual subtest scores in particular, and subsequently area scores, and Total Test scores may not be giving an accurate picture of a student's mathematics achievement. A difference of only a few points in Total

Test or area scores may have significant impact when a student's placement in a learning disability program is in question.

A study with an N of 40 carries considerable caveats for generalizing results, and the results of this study should not be a basis for discontinuing use of the Keymath-Revised, nor is that indicated. These results do however, indicate that caution should be used in placing emphasis on Keymath-R scores, particularly subtest scores. Examination of individual item responses on this test can yield a wealth of information about a student's specific mathematical abilities and the results of this study do not call this into question.

Recommendations

Further research should be done to determine whether the difficulty sequence on the Keymath-Revised test is significantly different for a referred population, than for the student population as a whole. This should be done with a sample that is large enough to allow generalizing of the results. It would be useful to study the difficulty sequence for a referred population on other commonly used mathematics achievement tests. These results raise similar questions about difficulty sequences on reading and other types of achievement tests.

The difficulty sequence presented in the test manual is not the same as the item administration sequence. The difficulty sequence was empirically determined after the test items had been put together in the test book. Although this administration sequence is constant for all individuals, encountering an item that is too difficult early in administration may make it more difficult to illicit good effort from some students on following items. It would be useful to conduct further research which looks at protocols of students from the normal population to determine how an out of sequence item effects the scores of individual students.

It is recommended that examiners use precaution when using Keymath-Revised scores to help make placement decisions, and that responses to specific items be the primary focus. The publishing of Keymath-Revised follows years of using the original Keymath and its wide acceptance as a diagnostic tool. It is important to remember that the the new version is not the equivalent of the original instrument, and it should be required to stand upon its own statistical soundness, and adherence to psychometric principles. Pending studies which support the use of Keymath-Revised, it should be used with the caution given to use of any new instrument.

REFERENCES

Bannatyne, A., (1973). Review of the Keymath Diagnostic Arithmetic Test. Journal of Learning Disabilities, 6(3), 131-132.

Connolly, A. (1988). Keymath-Revised (manual). Circle Pines, MN.: American Guidance Service.

Connally, A., Nachtman, W., & Pritchett, E. (1971). Keymath Diagnostic Arithmetic Test (manual). Circle Pines, MN.: American Guidance Service.

Greenstein, J. & Strain, P. (1977). The utility of the Keymath Diagnostic Arithmetic Test for adolescent learning disabled students. Psychology in the Schools, 14, 275-282.

Huebner, F., (1989). Test Reviews. Journal of Psychoeducational Assessment, 7, 364-367.

Kratochwill, T. R., & Demuth, D. M. (1976). Examination of the predictive validity of the Keymath Diagnostic Arithmetic Test and the Wide Range Achievement Test in exceptional children. Psychology in the Schools, 13, 404-406.

Lewis, R. (1989). Educational assessment of learning disabilities: A new generation of achievement measures. Learning Disabilities, 1(2), 49-55.

McCullough, B., Zaremba, B. (1979). Standardized achievement tests used with learning disabled and non-learning disabled adolescent boys. Learning Disability Quarterly, 2(4), 65-70.

Nicholson, C. (1988). New Tests. Council for Educational Diagnostic Services Newsletter, 2, 2-3.

Price, P. (1984). A comparitive study of the California Achievement Test (forms c and d) and the Keymath Diagnostic Arithmetic Test with secondary LH students. Journal of Learning Disabilities, 7, 392-396.

Satler, J. (1982). Assessment of children's intelligence and special abilities (2nd ed.) (P. 21). Boston: Allyn and Bacon

Sax, G. (1989). Principles of educational and psychological measurement and evaluation (3rd ed.) (p. 234). Belmont, Ca. : Wadsworth Publishing

Tinney, F. (1975). A comparison of the Keymath Diagnostic Arithmetic Test and the California Arithmetic Test used with learning disabled students. Journal of Learning Disabilities, 8, 313-315.

Appendix

TABLE I

RELATIVE POSITIONS OF DIFFICULTY FOR KEYMATH-REVISED TEST ITEMS

SUBTEST	ITEM SEQUENCE
Numeration	1 2 3 5 4 6 7 8 9 11 10 12 14 13 15 16 19 17 18 21 20 23 22 24
Rational Numbers	1 2 6 4 3 7 9 8 5 11 10 13 12 14 15 16 17 18
Geometry	1 3 2 4 5 7 6 8 9 10 12 11 15 13 14 16 17 18 19 21 20 22 23 24
Addition	1 2 3 4 5 6 7 8 9 10 11 12 13 14 16 15 17 18
Subtraction	1 4 5 3 7 2 6 8 9 10 11 12 13 14 16 15 17 18
Multiplication	3 2 1 4 5 8 7 6 9 11 10 12 13 14 15 16 17 18
Division	2 1 3 4 5 6 7 9 8 10 11 13 12 16 14 15 17 18
Mental Computation	1 2 4 3 6 5 7 9 10 8 12 14 13 15 11 16 17 18
Measurement	14 1 2 3 4 7 5 6 8 9 10 12 11 13 15 17 16 18 19 21 20 23 22 24

TABLE I (Continued)

Subtest	Item Sequence
Time and Money	1 3 4 2 6 5 8 9 7 10 12 16 14 13 17 11 15 18 20 21 19 22 23 24
Estimation	2 1 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
Interpreting Data	2 1 6 3 5 4 7 8 12 10 9 13 11 14 16 15 17 18
Problem Solving	2 1 3 4 5 6 7 8 9 10 11 12 13 14 15 16 18 17

TABLE II

ITEM DIFFICULTY SEQUENCE FROM STUDY SAMPLE

Subtest	Corresponding Difficulty Proportion Subtest Items Presented in Sequence										
Numeration	1.00	.975	.925	.90	.675	.625	.55	.525	.50	.375	
	1	8	10	14	15	17	16	18	19	20	
	2	9	13								
	3	11									
	4	12									
	5										
	6										
	7										
	.30	.175	.125	.025							
	21	23	22	24							
Rational Numbers	.85	.75	.50	.475	.45	.30	.225	.125	.10	.050	.00
	1	2	4	3	6	9	7	8	10	12	16
							5		11	13	16
										14	18
										15	
Geometry	1.00	.975	.95	.90	.85	.775	.725	.50	.475	.45	
	1	3	7	12	11	13	15	16	14	17	
	2	4	9								
	5	8									
	6	10									
	.30	.275	.15	.10	.075	.05					
	19	18	20	23	22	24					
	21										
Addition	1.00	.975	.925	.90	.875	.85	.825	.65	.175	.075	
	1	7	9	8	11	12	10	14	16	15	
	2					13				17	
	3										
	4										
	5										
	6										
	.050										
	18										

TABLE II (Continued)

Subtest	Corresponding Difficulty Proportion Subtest Items Presented in Sequence									
Subtraction	1.00	.975	.95	.85	.775	.725	.675	.575	.175	.15
	1	2 3 4 5	6	8	7 9	11	10 12	13	14	16
	.10	.025								
	15	17 18								
Multiplication	.95	.925	.90	.875	.85	.80	.725	.675	.625	.575
	2 4	1 3	7	6 8	5	9	10	12	11	14
	.50	.225	.20	.10	.075					
	13	16	15	17	18					
Division	.975	.925	.90	.80	.725	.70	.575	.475	.45	.375
	1	3	2	4	6	5 7 8	9	11	10	13
	.275	.20	.15	.10						
	12	16	14	15 17 18						
Mental Computation	.90	.875	.80	.725	.50	.475	.45	.35	.25	.225
	1	2	4	3	6	5 9	12	7 10	8	13
	.20	.10	.075	.05						
	14	11 15	16 17	18						
Measurement	1.00	.975	.95	.925	.85	.825	.775	.725	.625	
	1 2 4 5	3	6 7	8	9 10	12	14	11	15	
	.475	.375	.35	.25	.125	.075	.025			
	17	13	16	18 19	20 21	22	23 24			

TABLE II (Continued)

Subtest	Corresponding Difficulty Proportion Subtest Items Presented in Sequence									
Time And Money	1.00	.975	.875	.725	.70	.675	.65	.625	.425	.40
	2 3 6	1 4 5	7 8 9 11	12	10 16	13 14	17	15	20	18
	.35	.225	.175	.125	.075					
	21	19	22	23	24					
Estimation	.975	.925	.825	.75	.65	.50	.475	.30	.25	.225
	1	2 3	5 7	4	6	9	8	12	10	11
	.10	.075	.05	.025						
	14	15	13 17	16 18						
Interpreting Data	.975	.95	.925	.80	.75	.65	.55	.475	.425	.375
	1 2	3 6	5	4	9	7	8	10	11	13
	.35	.20	.125	.025						
	12	16	14 15	17 18						
Problem Solving	.95	.875	.675	.65	.60	.575	.475	.425	.325	.275
	1 2	3	8	4	5	6	9	10	11	13
	.25	.225	.125	.025	.00					
	7	12	14 15	17	16 18					

TABLE III

ITEMS WHICH VARY AT LEAST THREE POSITIONS BETWEEN STUDY SAMPLE AND NORMING SAMPLE

Subtest	Item Number	Sequence Position Presented in Manual	Sequence Position From Study Sample
Mental Computation	12	11	8
Time and Money	1	1	4
	11	16	10
Interpreting Data	9	11	7
Problem Solving	7	7	12
	8	8	4
Geometry	3	2	5
	6	7	4
	7	6	9
Addition	10	10	13
Subtraction	7	5	8
Multiplication	5	5	8

2

VITA

Dan O. Hooker

Candidate for the Degree of

Master of Science

Thesis: COMPARING DIFFICULTY SEQUENCE OF KEYMATH-REVISED TEST ITEMS FOR THE NORMING SAMPLE AND FOR STUDENTS REFERRED FOR LEARNING DISABILITIES ASSESSMENT

Major Field: Applied Behavioral Studies

Biographical:

Personal Data: Born in Austin, Texas, August 30, 1954, the son of Edward W. and Marjory P. Hooker.

Education: Graduated from Thomas Edison Senior High School, Tulsa, Oklahoma, in May 1973; received Bachelor of Science Degree in Sociology from Oklahoma State University in December, 1980; completed the requirements for the Master of Science degree at Oklahoma State University in May 1990.

Professional Experience: Psychometrist/Prescriptive Teacher, Child Service Demonstration Center, August, 1988 to May, 1990.