



National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

UNIVERSITY OF ALBERTA

A VALIDATION OF THE MILLER ASSESSMENT FOR PRESCHOOLERS

BY

SARA YARWOOD

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH IN PARTIAL
FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF EDUCATION

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA

(FALL 1991)



National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-70266-4

Canada

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Sara Yarwood

TITLE OF THESIS: A Validation of the Miller Assessment for Preschoolers

DEGREE: Master of Education

YEAR THIS DEGREE GRANTED: 1991

PERMISSION IS HEREBY GRANTED TO THE UNIVERSITY OF ALBERTA LIBRARY TO REPRODUCE SINGLE COPIES OF THIS THESIS AND TO LEND OR SELL SUCH COPIES FOR PRIVATE, SCHOLARLY OR SCIENTIFIC RESEARCH PURPOSES ONLY.

THE AUTHOR RESERVES OTHER PUBLICATION RIGHTS, AND NEITHER THE THESIS NOR EXTENSIVE EXTRACTS FROM IT MAY BE PRINTED OR OTHERWISE REPRODUCED WITHOUT THE AUTHOR'S WRITTEN PERMISSION.



SARA YARWOOD

87 STERLING STREET

HAMILTON

ONTARIO

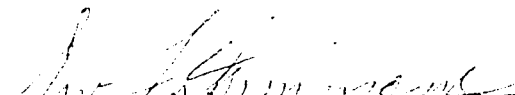
CANADA

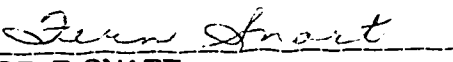
DATE: October 2, 1991

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

THE UNDERSIGNED CERTIFY THAT THEY HAVE READ, AND RECOMMEND TO THE
FACULTY OF GRADUATE STUDIES AND RESEARCH FOR ACCEPTANCE, A THESIS
ENTITLED: A VALIDATION OF THE MILLER ASSESSMENT FOR PRESCHOOLERS
SUBMITTED BY: SARA YARWOOD
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF: MASTER OF EDUCATION.


DR. G. FITZSIMMONS (SUPERVISOR)


DR. F. SNART


DR. F. PETERS

DATE: October 3, 1991

I dedicate this thesis to the two people who, more than any others, have changed my life over the time I have spent in Edmonton, and to whom I cannot adequately express my admiration, warmth, and affection. Graham Ross and Patricia Schuster have been the best of friends, the best of colleagues, and the best of employers. Their contributions to my personal and to my professional growth and development have been both nurturing and loving. It has been my privilege to be connected with them.

ABSTRACT

The Miller Assessment for Preschoolers was administered to all children entering the E.C.S. program in St. Albert School District #3 in the spring prior to their school entry (May, 1989). Results of that screening program and results of standardized assessment conducted during those children's grade one year (1990/91) were amalgamated in order to determine to what extent scores on the Miller Assessment for Preschoolers were correlated with achievement in grade one. Standardized assessment follow up measures included the Gates Reading Readiness Test, the Canadian Test of Basic Skills, and the Gates MacGinitie Reading Tests.

Correlations between two forms of the Miller Assessment for Preschoolers and the three outcome measures were calculated. Generally correlations between MAP scores and achievement scores were weak to moderate. In some instances the correlations did not reach statistical significance. These results provide no support for the validity of the MAP as a predictor of achievement as measured by standardized tests administered throughout the grade one year. Implications for the school's screening program include the need for careful reconsideration of the use of the MAP for this purpose. Practical implications relate to the need for carefully considered allocation of screening and assessment resources. Implications for further research include the need to develop and evaluate ecologically based models of predicting achievement.

ACKNOWLEDGEMENTS

There are many people whose support and encouragement during the past four years have been of immeasurable value. People whom I've met in the Department of Educational Psychology have come and gone, whilst I've lingered on, yet many have remained as true friends and some have even been around as long as I have! I'd particularly like to thank Rod Beattie, Shawn Crawford, Elizabeth Starr, and Chris Wilson who have all exhibited confidence in me (not to mention patience).

The next group of people who must be acknowledged, come second only because they are associated with the second stage of completing this degree. The people that I have worked with professionally over the past two years have, in many ways, taught me more than my academic studies. I've developed confidence; I've learned to be fragile at times, in order to be strong; and I've experienced how wonderful it is to work with wonderful people. Pat Schuster and Graham Ross are the driving force behind the team. Shawn Crawford, Greg Erikson, and Marlis Kozak-Krueger are also special people with whom I consider it a privilege to work. Everyone who works with E.C.S.S. deserves special recognition for the support and care which they give to each other.

My family has not been here in Edmonton as I've worked on this Master's degree, but they've been behind me, and in their love and support have helped me to get to where I am. Many thanks to each of them.

TABLE OF CONTENTS

	Page
CHAPTER 1 INTRODUCTION	1
CHAPTER 2 REVIEW OF THE LITERATURE	
Introduction	5
Test Development and Structure	5
Psychometric Characteristics	8
Research and Comments Establishing Validity	12
Reviews and Research Using the MAP	17
Problems with the Literature	21
Objectives	22
Statement of Hypotheses	22
CHAPTER 3 METHODS AND PROCEDURES	
Introduction	24
Nature of the Study	24
Sample	25
Instruments	29
Data Collection and Recording	37
Data Analysis	39

CHAPTER 4	RESULTS	
	Introduction	4 1
	Gates MacGinitie Reading Readiness Rest	4 3
	Canadian Test of Basic Skills	4 7
	Gates MacGinitie Reading Tests	5 0
	Summary	5 2
CHAPTER 5	DISCUSSION	
	Introduction	5 4
	Discussion	5 4
	Limitations	6 1
	Implications	6 3
	Preschool Screening Models and Practice	6 7
	Recommendations	7 1
	Implications for Further Research	7 3
	Summary	7 3
	REFERENCES	7 5

I. Introduction

For the past two years a Roman Catholic suburban school district in Alberta, Canada has been administering the Miller Assessment for Preschoolers (MAP) to all children entering its Early Childhood Services (E.C.S.) program. According to a counsellor at one of the schools where the MAP is administered, the purpose of using this instrument has been to identify children with previously undiagnosed developmental delays (personal communication). Additionally, it has been the district's goal to use the time taken to administer the MAP as an opportunity to familiarize themselves with the children entering their program, in order that appropriate classroom placements might be made. This is particularly important as the available E.C.S. programs include both regular kindergarten programs, and "cluster programs" designed to meet special needs.

Historically, it had been the policy in the district to administer the Peabody Picture Vocabulary Test - Revised (PPVT-R), the Beery Developmental Test of Visual Motor Integration (VMI), and an informal language assessment to each child entering kindergarten. Two years ago it was decided that this screening process should be changed in order to use a single standardized instrument that would be more predictive of academic difficulties. Several assessment instruments were reviewed, and the MAP was chosen as it appeared to best address the system's needs. The use of the MAP is intended to identify children who may develop academic problems.

Due to lack of information concerning progress in school for the children screened using this instrument, a problem currently exists in assessing the adequacy of the MAP for fulfilling its intended purpose. The MAP was chosen in

the hope that it would be useful in predicting academic problems. Evaluation of its use relies, in part, on obtaining an objective measure that would describe the extent to which the MAP fulfills this purpose.

This research is intended to aid in determining if MAP scores are correlated with achievement scores in order to evaluate that aspect of the information obtained from the administration of the MAP which is directly pertinent to its use as a screening instrument in this district. Predictive validity is the main issue that will be addressed.

The legitimacy of the use of the MAP as a screening instrument in this district, can be evaluated, in part, through determining if children who do well on the MAP subsequently maintain adequate academic achievement. Concomitantly, it is necessary to determine whether children with low scores on the MAP do have a greater likelihood of experiencing academic delays. As well, information with respect to the adequacy of the published norms for this district will aid in more precise identification of the children to be served in special needs programs.

If the Miller is fulfilling its purpose, then scores on the MAP should predict scores on subsequent achievement measures at better than chance levels. That is, low scores on the MAP are expected to be associated with lower achievement (as defined by scores on the Gates MacGinitie Reading Readiness Test, the Canadian Test of Basic Skills, and the Gates MacGinitie Reading Test), and high MAP scores should be associated with at least average achievement in grade one. The MAP assesses a wide range of developmental tasks. Any differences found in its ability

to predict scores across the outcome measures may provide information that is relevant to interpreting the constructs that it measures.

Chapter Two of this thesis contains a review of the literature pertaining to the MAP, and focuses upon its psychometric properties and predictive validity. For the most part, research conducted with the MAP is described in the revised manual (Miller, 1988). This research is reviewed here, as is independent research conducted using the MAP. Test reviews, and comments from other sources are also included. The literature review is concluded with a statement of the hypotheses driving the current study.

Chapter Three of the thesis describes the methods and procedures of the current investigation. This chapter includes an Introduction and a description of the Nature of the Study. These are followed by descriptions of the Sample, Instruments, Data Collection, and the Data Analysis procedures. The section entitled Nature of the Research describes both the archival nature and form of the data. The section describing the Sample includes a statement explaining the inclusion criteria for both subjects and data. The Instruments section includes descriptions of the major standardized measures. This section also describes the school registration form, which was used to confirm demographic information such as date of birth and health history. The section describing the Data Analysis provides statistical statements of the hypotheses to be tested, and indicates which analysis techniques were used. Where appropriate, comments relating to the choice of techniques are also included.

Chapter Four of the thesis provides the results of the analyses conducted. This information is presented in a tabular or graphic form where appropriate.

Verbal descriptions of the quantitative results are provided, and integrated to some extent with interpretation of their meaning and implication. More general or theoretical commentaries with respect to interpretation of results are reserved for Chapter Five.

Chapter Five completes the text of the thesis. It includes a summary and discussion of the results, and integrates the results with the issues raised in the literature review. The limitations of the present study are then described. Conclusions are then drawn, and the text ends by providing directions for further research. A list of the References used, and the complete appendices are found after Chapter Five.

The goal of the research has been to provide: (a) a brief description of the degree to which the norms of the MAP are adequate for the district's programs, (b) quantification of the extent to which the MAP is predicting academic achievement in the first year of school, and (c) an indication of whether the MAP differentially predicts global achievement or areas of achievement. Thus, the importance of the study revolves around its ability to provide information that could be used by the school district as part of an evaluation of current screening policy. Any such evaluation undertaken by the district would be likely to involve a broader scope of consideration than can be included in a study of this nature. Therefore, it is expected that the results ensuing from this study would be incorporated with a variety of other evaluation parameters. This study is limited to those described above.

II. Review of the Literature

Introduction

This chapter presents a review of the literature pertaining to the Miller Assessment for Preschoolers. It includes a review of information pertaining to the development, norming, and psychometric properties of the MAP, as well as a review of the research documenting its predictive validity. The problems in the literature are described and the objectives and hypotheses driving the current study are then presented.

Test Development and Structure

The MAP is described in its revised manual as an instrument designed to assess developmental status of children aged two years nine months to five years eight months, across a range of content. This content includes the motor, cognitive, and behavioural domains. The goals of the test developer, as stated in the manual, were to provide a statistically sound assessment procedure to identify developmentally delayed children, as well as to provide a clinical framework for identifying areas of strength and weakness and providing possible avenues for remediation (Miller, 1988).

According to the manual (Miller, 1988) the MAP was developed over a period of approximately ten years, during which time thousands of children were assessed and hundreds of items were field tested. After the field tests, which reduced the number of items, a tryout edition of the MAP was used to reduce the items still further, to the final 27 included in the current version. Five hundred and thirty items were actually field tested with a nationwide sample (Miller, 1984).

The MAP yields a total score in the form of a percentile as well as scores on five Performance Indices. The Performance Indices are purported to assess three types of developmental abilities. These include Sensory and Motor (measured by the Foundations and Coordination Indices, having 10 and 7 items respectively), Cognitive (measured by the Verbal and Non-Verbal Indices, having 4 and 5 items respectively), and Combined (measured by the Complex Tasks Index which is comprised of 4 tasks). Some items are represented in more than one Performance Index. Scores on all 27 items also yield a Total score which is provided as a percentile at age.

Foundations

Miller (1988) claims that the Foundations Index measures neurological and neuromaturational facets of development. Included in these categories are measures which attempt to reflect the child's sense of position and movement, sense of touch, and the development of movements patterns. The author contends that "Poor performance on items which sample neurological abilities has been found to be a significant predictor of later learning difficulties." (Miller, 1988, p. 84). However, Miller does not report the magnitude of such predictive coefficients, nor the relative value of such measures over more direct measures of early learning.

The Foundations Index is made up of ten items which attempt to reflect developmental maturity. Examples of items on the Foundations Index include the Romberg item which requires the child to stand with feet together and maintain balance (with eyes closed) for fifteen seconds, the Hand-to-Nose item, which reflects the child's ability to move the hand in a straight and accurate line while

eyes are closed, and the stereognosis item which requires the child to identify a shape placed in his or her hand through the sense of touch alone. Many items are similar to those encountered on neurological screens such as the Quick Neurological Screening Test (Academic Therapy, 1987)

Coordination

This index purports to measure the physical coordination of motor movements. Gross and fine motor coordination, as well as oral motor abilities are measured. Skills include tongue movements, rapid alternating movements (which requires the child to stamp feet in an alternating pattern), and motor accuracy (which requires the child to draw a number of short, straight vertical lines). Again, Miller argues that research has demonstrated that children with learning problems exhibit deficiencies in these skills. However, although she refers to empirical evidence for these claims, she does not report specific statistics supporting these assertions, nor does she describe methodology leading to these conclusions.

Verbal Index

The Verbal index of the MAP purports to reflect language skills which provide a measure of language functioning. Only four items are included on this index. These include a general information item, requiring the student to answer three questions, sentence repetition and digit repetition tasks which measure short term memory, and a task which requires the child to follow directions. Despite the fact that Miller cites researchers who state that language impairments comprise the largest subgroup of learning disabilities, less than twenty percent of MAP items are included on this scale.

Non-Verbal Index

Miller states that research into learning disabilities consistently supports the connection between non-verbal cognitive dysfunction and school learning problems. Five non-verbal items are included on this index. They include measures of visual memory, sequencing, and figure-ground relationships.

Complex Tasks

The complex tasks index consists of items considered by the test author to reflect a combination of skills. These include puzzles, mazes, and block designs.

Psychometric Characteristics of the MAP

Reliability and validity of the MAP are discussed in the manual and have also been reviewed by a number of independent sources. Miller (1988) reports that the interrater reliability, the internal reliability, and the test-retest reliability of the MAP have all been determined. Interrater reliabilities were Pearson product-moment correlations between pairs of scores obtained by field supervisors who each tested half of a sample of 20 children and observed the other half being assessed. Interrater reliabilities for each Performance Index fell above .97 with the exception of the Coordination Index, the reliability of which fell at .84. The reliability of the Total MAP was presented as .98. The author conceded that due to methodological aspect of this study the reported reliabilities may be somewhat inflated. It should also be noted that the sample size involved was relatively small (N=40, randomly selected from the standardization sample), and that the examiners in this case were field supervisors in the standardization of the MAP. Further investigations of results obtained by examiners with less extensive training and experience, might

provide more accurate estimates of inter-rater reliability under more realistic field conditions.

Test-retest reliability was established using a sample of 81 children originally included in the standardization sample. The two administrations took place 1 to 4 weeks apart. Stability ratios were described as percentages of children whose scores on the Total MAP and on each of the Performance Indices remained the same over the two test administration. Retest stability for the Total MAP was 81%, with Indices' stability ratios ranging from 72% to 91%. Unfortunately, no other descriptions of the test-retest reliability were provided. It is unclear as to why Miller chose to report test-retest reliabilities in this manner. More traditional test-retest reliability studies, providing coefficients which would be comparable to those provided by other screening measures, would allow the consumer to make more informed decisions regarding reliability.

Miller states that it can be concluded that the MAP "is quite stable over time" (1988, p. 105). However, she does not provide descriptions of psychometric standards provided by measurement experts nor directly compare the MAP's characteristics to any such standards in support of this claim.

Internal consistency (referred to by Miller as internal reliability) was obtained for the Total MAP norming sample by employing the split-half method and using the Spearman-Brown correction formula. A reliability of .79 was obtained. Miller does not explain exactly how these reliabilities were calculated. For example, although she states that reliability estimates were based on "raw scores" she does not indicate whether items were scored dichotomously,

trichotomously, or on the basis of the exact number of "sub-items" within each item.

Sattler (1988) states that "for most tests of cognitive and special abilities, a reliability coefficient of .80 or higher is generally considered to be acceptable" (p.25). Sattler also notes that "high reliabilities are especially needed for tests used for individual assessment." (p. 25). However, Sattler does not indicate which estimate of reliability must meet the .80 criteria. In that discussion it is not stated what is meant by "high" reliabilities that individual assessment instruments should meet.

Bracken (1987) proposes criteria specifically for preschool measures. He states that subtests' internal consistencies should be .80 or greater, total test internal consistency should be .90 or better, and total test stability (test-retest reliability) should reach .90. Criteria for acceptability must be related to the meaning of reliability with respect to measurement error. Nunnally (1970) describes this succinctly:

Measurement error, or unreliability, *always* (italics added) work to obscure, or as we say, *attenuate* any type of scientific lawfulness. Whatever *real* lawfulness there is in nature will appear blurred if relatively unreliable measures are used to chart that lawfulness. When dealing with predictor tests, this means that to the extent to which the test has measurement error, it cannot do a good job of predicting a criteria. (p.199).

Nunnally presents similar arguments with respect to multiple sources of error. To the extent that a test score reflects multiple sources of unreliability or error, it cannot do a good job of predicting criteria. If the correlation is

attenuated, the prediction equation (which is directly related to the correlation coefficient) must also be attenuated. As well, it is true that to the extent to which the criterion (or outcome) measure is unreliable, the prediction equation will be attenuated. Both sources of unreliability will affect the relationship, and therefore the equation describing that relationship, between a predictor and criteria. Therefore, in order to optimize the ability to identify any "real lawfulness", by determining the equation which describes such a relationship, it is important to minimize error from both sources.

Miller (1988) also reported that an average item-to-test correlation was determined, and this too was subjected to the Spearman-Brown correction formula (which increases the reliability estimate). A reliability estimate of .82 was obtained. The author claims that due to the heterogeneous nature of the tasks, these reliability estimates are psychometrically adequate. This is debatable. Especially with respect to average item to test correlations, it is reasonable to expect higher coefficients, since the average .82 suggests that either few items reach much higher correlations to the total test score or that some items have item-to-test correlations substantially lower than .82, raising questions as to their usefulness as part of the overall measure.

No internal consistencies for Indices were reported due to the small number of items within them. Unfortunately, Miller does not address the issue as to whether such indices should be reported at all, if they contain too few items to allow for internal consistency estimation. Further difficulties with respect to the index scores reported are addressed in the section of this chapter which discusses the validity of the MAP.

Research and Comments Establishing Validity

Miller (1988) documents evidence in support of content, criterion-related (including concurrent and predictive), and construct related validity. The relationship of MAP items to theory is supported through a review of the supporting literature and through a procedure involving expert ratings of the content validity of the items considered for inclusion. Content specifications by item are also provided.

Concurrent validity was investigated in part through comparisons with other standardized instruments. These included the Wechsler Preschool and Primary Scale of Intelligence (WPPSI), the Illinois Test of Psycholinguistic Abilities (ITPA), the Southern California Sensory Integration Tests (SCSIT) and the Denver Developmental Screening Test (DDST). With respect to the first three instruments, correlations are provided between subtests and/or indices as well as total test scores. Most reported correlations were weak (e.g. MAP with WPPSI $r=.21$ which was not statistically significant), yet Miller interprets evidence of validity on the basis of statistical significance of some of the reported correlations. No correlations on any of the concurrent validity measures were greater than an absolute value of .45. Even the largest coefficient indicates that only twenty percent of variance in scores is shared between measures. In fact, the evidence supplied in the manual for the MAP seems to refute Miller's claims of concurrent validity of the MAP with the other measures investigated.

It should also be noted that Miller advises caution in interpreting these results due to the small sample sizes used ($N's=30$). As the samples are small, Miller seems to be suggesting that larger samples might have yielded statistically

significant coefficients. While similar coefficients might well be significant if the size of the sample was substantially increased, it does not follow that the evidence for concurrent validity would be greater. Larger sample sizes would not in anyway guarantee a greater magnitude of correlation. Thus, if the reported correlations are in any way accurate reflections of the actual relationship of the test scores, the variance between the measures overlaps minimally. This would not change were the coefficients to reach statistical significance purely on the basis of larger samples, since larger samples only require a coefficient of less magnitude in order to be confident that the coefficient was not different from zero due to chance alone.

With respect to the DDST which is not a standardized measure, Miller reports the percent of examinees whose categories differed on the relatively similar categories provided between the two instruments. 28% of the sample (n=30) were grouped differently. The MAP placed many more children in at-risk categories, which included Yellow (watch carefully) and Red (refer). Miller suggests that this study should be interpreted cautiously as it does not necessarily indicate that the Miller is more accurate at identifying children for future problems. In fact, it may provide evidence that the MAP over identifies in terms of percentage. That is, since the cutoffs were chosen to identify 25% of children in the "watch carefully range" Miller may have chosen a cut off score which identifies too large a proportion of the population. It seems likely that most schools would find comprehensive diagnostic assessment of 25% of its incoming E.C.S. population to be financially prohibitive. Miller does not explicitly state that such comprehensive follow up is necessary with all children

who fall in the "Yellow" or "watch carefully" range, but does not provide alternatives which provide a practical policy for using MAP results.

Predictive validity studies have, for the most part, been conducted by the author of the MAP. As they were obtained Miller published these results in professional journals. Summaries of research completed prior to publication, were published in the 1988 revision of the MAP manual. The predictive validity studies have mainly been based on a four year follow up of 338 children included in the original standardization of the MAP. Eight standardized assessment instruments were administered. Ratings on six school performance criteria were obtained. Correlations with standardized instruments ranged from a low of .21 with the Developmental Test of Visual Motor Integration (Beery) to a high of .50 with the Full Scale Wechsler Intelligence Scale for Children - Revised (WISC-R) score. Correlations with academic achievement measures (Woodcock Johnson Psychoeducational Battery (WJB)) were .36, .38, and .35 respectively for Reading, Math, and Written Language. Correlations with School Performance Criteria ranged from a low of .17 with report card math grades to a high of .24 with special class placement and report card language grades. Despite these low correlations, a t-test comparing the mean Total MAP scores for "Problem" and "No-problem" children on each of the six school related criteria indicated that the problem group obtained significantly lower Total MAP scores.

Miller argues that the statistically significant correlations between MAP results and outcome measures provide evidence of validity. In fact, even the highest of these correlation (.50) indicates that only 25 percent of the variance amongst scores is shared between these measures. In terms of the practical

usefulness of this degree of correlation for predicting academic performance on the basis of MAP scores, there is little to recommend such a procedure on the basis of these predictive validity coefficients.

Additional examinations of the MAP's predictive validity were conducted by detailing classification accuracy between the MAP Total score and the outcome measures, using both the 5th and 25th age related percentile cut-off points suggested for the MAP. Overall Percent agreements with the outcome measures ranged from 82.25% to 92.60% using the 5th percentile cut-offs, and from 73.37% to 78.11% when using the 25th percentile cut-off.

Miller attempted to demonstrate construct validity of the MAP through factor analysis, item-age trends, correlations amongst items, correlations between indices and total score, and by examining the rate of correctly identifying pre-categorized problem children. Although Miller states that there is evidence for factorial validity of the subgroupings (indices), in fact the factor analysis indicates that some items from three of the Indices (Coordination, Verbal, Non-verbal) group together, while the other indices do not show patterns of items loading on corresponding factors.

The Factor Analysis which Miller reports (1988) yielded six factors, in contrast to the five indices which are used in scoring the MAP. Although validity for the three indices mentioned may have been obtained, Miller concedes that the Foundations Index and the Complex Tasks Index items do not appear to measure a single domain of behaviour. Rather, they seem to represent a variety of skills or developmental tasks. As Miller notes, this suggests that for these two Indices at least, no singular construct has been identified or validated. As for the three

factors identified which did not match any of Miller's performance indices, Miller makes no attempt to interpret their structures or the constructs reflected. Additionally, it should be noted that many of the factor loadings reported were below .50, suggesting that the factors identified with the Varimax rotations are not being measured in a refined manner by the MAP items which do load on them. The exceptions are the factor loadings for three items on the Verbal factor which load .61, .61, and .69 respectively. Interestingly however, the "block designs" item also loads on this factor, suggesting that a factor more similar to a general intellectual factor may be the underlying dimension.

Miller describes the construct validity as being supported by the item age trends evidenced on several items of the MAP. Miller states that each MAP item discriminates between various age levels. It should be noted however, that Miller is not clear as to how this provides evidence of construct related validity. The scoring rules were based on the abilities of children at different ages to complete task requirements. If Miller is stating that because children improve with age there is evidence that a specific developmental construct is being measured, then her argument is based on circular reasoning if in fact she designed scoring criteria solely on this basis, rather than on task analysis or task characteristics.

Correlations between items and Total MAP scores ranged from a low of .20 to a high of .50. Performance Index scores were also correlated with Total MAP scores. These correlations ranged from a low of .65 to a high of .78. In a study of MAP scores of children already identified as having pre-academic problems, 75% of children obtained Total scores in either the Yellow (watch carefully)

range, or the Red (definite concern) range. When the top four age groups were considered apart from the two youngest groups, the rate rose to 84%. Although Miller describes this as evidence of the MAP's validity, careful evaluation of these figures reveals that even if a district referred or programmed for all children below the 25th percentile on the MAP, one in four children who have pre-academic problems would be missed.

According to Miller "Studies of the MAP indicate that this instrument is both reliable and valid. Examiners should feel confident that the MAP Total score is an accurate measure of a child's true level of ability." (1988, p.123). However, these assertions appear dubious at best, when the evidence presented is examined from a different perspective than that of the test's author.

Reviews and Research using the MAP

A search of published literature (to June 1991) indicated that little research regarding the use of the MAP has been published in professional journals. Miller is associated with each of the research articles relating to data with respect to predictive validity of the MAP. These studies have also been reported in the 1988 revision of the manual for the MAP and have been described above.

The most recent reference to the MAP in a search of psychological literature was to a Canadian study conducted by Daniels and Bressler in 1990. These authors acknowledged the lack of studies additional to those required for test publication. In the literature review they refer to two unpublished research studies which "provide preliminary support for the MAP's ability to differentiate between academic 'problem' and 'no problem' children and support for the 5% and 25% cutoff points." (sic) (Cohn, 1986, and Lemerand, 1985

cited in Daniels and Bressler, 1980, p. 49) Daniels and Bressler discuss the need for validity studies with referred populations as they note that the MAP is in wide use among occupational therapists in British Columbia. They conducted a study analyzing MAP results of 95 subjects referred to the Occupational therapy department in a Children's hospital. Subjects were grouped according to diagnosis including, speech-language problems, output problems, speech-language output problems, developmental delay, psychiatric diagnosis, or no diagnosis.

One way ANOVA's were performed on total and index scores (in the form of percentiles) and mean differences between groups were noted. The authors provide evidence which suggests that MAP scores do differentiate between the diagnosed and undiagnosed groups. The authors state that this provides preliminary evidence in support of the use of the MAP with developmentally delayed referred populations. However, they are cautious in their interpretation and note that further research is needed before professionals can be confident of the validity of using the MAP in a clinical setting.

It should also be noted that the authors only report that the children were referred, demographic information was recorded and the children were routinely given the MAP. They indicate that the child's "medical/developmental diagnosis" was also recorded. However, it is unclear from the authors' report if in fact the diagnosis was completely independent of the results of the MAP (e.g. made by a clinician unfamiliar with the MAP results for the child). If the diagnosis was made in part on the basis of MAP scores and/or patterns of scores, group differences amongst scores and patterns would be expected, and the results may

provide minimal added information with respect to the validity of the use of the MAP for making such diagnoses or any other use with the population to which these authors intended to generalize their results.

Four review articles which discuss standards for screening instruments included descriptions of the MAP, and provided critical analysis of its use as an assessment instrument in various capacities. Padgett (1985) reviewed the MAP, describing its standardization, reliability and validity, as well as the organization, materials and administration, scoring, and interpretation guidelines. Padgett concluded that "While the MAP consists of well designed and carefully standardized items, its overwhelming preponderance of visual-motor tasks greatly minimizes its usefulness in screening for possible learning problems." (1985, p. 188). This conclusion is based on the author's argument that current research indicates that most children who experience academic delays or learning disabilities are affected by language based problems and not by visual-motor problems. Thus, this author presents a theoretical argument against the use of the MAP but does not provide research evidence that the MAP is not useful in identifying pre-academic problems.

In contrast to the opinion of Padgett (1985), Slaton (1985) suggests that "Used appropriately and interpreted with caution the Miller Assessment for Preschoolers may prove to be our most valuable tool for identifying children with pre-academic or school related problems." (p.70). This conclusion appears to be based on a review of the psychometric properties of the MAP and the assertion that the MAP identifies children as being at risk who tend to do well on traditional developmental assessments. These traditional developmental

assessments have failed in the past to be sensitive to mild or moderate disabilities. No data to support Slaton's (1985) assertion is provided, although it is assumed this is an interpretation of the reliability and validity statistics provided in the test manual. Slaton does not engage in a critical analysis of the psychometric properties of the MAP, but rather appears to accept Miller's interpretations of the MAP's psychometric characteristics.

Bracken (1987) reviewed ten preschool assessment instruments using seven main psychometric criteria. Criticisms with respect to the MAP included lack of subtest reliabilities, inadequate total test reliability (when compared to the .90 level expected for diagnostic purposes and .80 expected for screening purposes), and the lack of available data to evaluate many of the other psychometric criteria. Overall, Bracken's analysis allows the reader a new perspective and interpretation of the MAP's psychometric properties. It suggests much greater caution with respect to the use of the MAP than that provided either by Miller or some other reviewers.

Scores on the Miller Assessment for Preschoolers (MAP) have been compared with those on the Peabody Developmental Motor Scales (PDMS) by Provost, Harris, Ross, and Michnal (1989). Weak to moderate correlations were obtained. Although the PDMS Fine Motor Scale identified the most children as delayed, the MAP identified some children as delayed who were not identified by the PDMS (and vice versa). The authors suggest that they have obtained support for theories which state that appropriate tactile abilities (as measured by the MAP) are necessary for normal fine motor development. Since the authors

expect these skills to be important for school success, there is some support for the content validity of the MAP as a preschool screening measure.

Problems in the Literature

The main problems in the literature are with respect to the dearth of published research in which the test's author is not involved, especially with respect to predictive validity. As well, the absence of Canadian studies (to date) limits the available knowledge upon which to base the use of the MAP in a Canadian population. It appears that no Canadian studies which have addressed the use of the MAP have been published in research journals. No Canadian norms are available and it is not possible to determine without local studies the validity of using this instrument. Although there have been some results which could be considered encouraging with respect to predictive validity, these results need to be replicated locally if the school district is to be confident about the use of this instrument. In addition to these problems with research conducted *since* the publication of the MAP, there are a number of problems with the reliability and validity information reported in the portion of test manual which describes the development and standardization of the instrument. The discussion included here will also address these issues.

The current investigation was aimed at addressing the need for the school district in question to determine if scores on the the Miller Assessment for Preschooler's are related to future academic success. Since the main goals of the screening procedure are to identify children who would be predicted to experience academic difficulty, the extent to which scores on the MAP correlate

with scores on academic achievement instruments, is the main question of interest. Thus, the investigation had the following objectives.

Objectives

The objectives of the investigation were to:

1. Determine if there are significant correlations between scores on the Miller Assessment for Preschoolers (administered prior to kindergarten entrance) and scores on academic readiness and achievement measures (obtained in grade one).
2. Determine if lower scores on the MAP are differentially associated with outcome measure scores. That is, the objective is to determine if there is a different correlation with outcome measure scores when subgroups of MAP scores (e.g., low, middle, and high) are considered.

These objectives lead to the following research questions

Statement of Research Questions

1. Is there a significant correlation between scores on the Miller Assessment for Preschooler's administered prior to kindergarten entrance and scores on the Gates MacGinitie Reading Readiness measure administered at the beginning of grade one?
2. Is there a significant correlation between scores on the Miller Assessment for Preschooler's administered prior to kindergarten entrance and scores on the Canadian Test of Basic Skills administered at in the middle of grade one?
3. Is there a significant correlation between scores on the Miller Assessment for Preschooler's administered prior to kindergarten entrance and scores on the Gates Reading Tests administered at the end of grade one?

The Miller has a positively skewed distribution and is not intended to discriminate between average and above average subjects. Therefore, it seems possible that even if the MAP scores do not correlate with outcome measure scores for the group as a whole, scores for children scoring in the low range on the MAP may correlate more strongly with scores on the outcome measures, than they will for children scoring in the middle or high range on the MAP. This conjecture leads to the following further research questions.

4. Do low scores on the MAP correlate more highly with Gates MacGinitie Reading Readiness Test scores than middle range or high scores?
5. Do low scores on the MAP correlate more highly with Canadian Test of Basic Skills scores than middle range or high scores?
6. Do Low scores on the MAP correlate more highly with Gates MacGinitie Reading Test scores than middle range or high scores?

III. Methods and Procedures

Introduction

This section will outline the components of the current study including the nature of the study, sample, instruments, data collection, and analysis procedures.

Nature of the Study

The current investigation is based on the amalgamation of archival data already collected by the school district for other purposes. Thus, it involves a retrospective view of data whose analysis was not preplanned for the purpose of this research. The limitations and benefits resulting from the design are discussed in chapter five of this thesis. The investigator was interested in pursuing analysis of the available data that would provide information with respect to the goals described in chapter one. That is, information which would help in evaluating the degree to which the MAP is fulfilling its stated purpose.

Permission was granted directly from the school district to access the data kept in student records. Since the data were archival in nature, was already held by the district, and was to be amalgamated in a manner which ensured anonymity of the children in question, only the consent of the school district (through the director of student services) was necessary.

The data were amalgamated in two stages. Data from the Miller Assessment for Preschoolers were obtained by reviewing records for student's enrolled in grade one during the 1990/91 school year. Data from all achievement measures were obtained by reviewing the same school records at the end of the grade one year, when all such results had been recorded.

Given the nature of the data, and the manner in which they were amalgamated, the research design of the current investigation is purely correlational. The degree of association between scores on the standardized measures in question is the major factor of interest. No causative generalizations are implied, nor intended. The sampling method allows for some generalizations to be made with respect to children typically included in the school district in question. Generalizations to any other population are tentative, and should only be made with due consideration and caution.

Sample

Selection Procedure

All parents registering children in Early Childhood Services (ECS) programs in the 1989/90 school year were informed that a screening procedure was being instituted, and their participation was requested. Parents brought their children to the school in May or June of 1989 and the MAP was administered. Some students not screened during May or June were administered the MAP in early September. Thus, the vast majority of children enrolled in Kindergarten were included in the screening procedure. Subsequent to the MAP administration children were enrolled in E.C.S. programs in September of 1989. Six French Immersion programs and 10 English programs (including the "cluster programs") were run from September 1989 to June of 1990.

Description

The children included in this screening procedure ranged in age from approximately four years three months to five years three months at the time of the initial screening procedure. Thus, two levels of the MAP were administered.

These were the four years three months to four years eight months form (MAP IV), and the four year nine months to five year two months form (MAP V). As only these two forms were used, there were occasional occurrences of a child older or younger than these ages who was administered the form most closely matching his or her age. In these cases the subject was excluded from the data analysis. Thus, two groups of children were examined.

It is likely that the sample represents a variety of socioeconomic conditions reflecting distribution of income and occupations typical for an middle to upper middle class suburban community. However, confirmation of this supposition could not be obtained from the available information. Of the children for whom MAP scores were analysed for the purpose of this study, 83.2% of children had at least one parent who stated on the school registration form that they are Roman Catholic. 16.8% of mothers reported other Christian affiliations, while 8.1% of mothers reported that they had no religious affiliations or left the item blank. Three of the mothers (1.5%) indicated other religious affiliations. 22.3% of fathers reported other Christian affiliations, while 19.3% reported that they had no religious affiliation or left the item blank. One father reported a non-Christian faith, accounting for approximately .5% of the sample of children on whom the MAP data were analysed.

Although it was expected that the distribution of males and females would be approximately equal, 55% of the sample on whom MAP data were collected was male, and 45% were female. No explanations for this disproportion was forthcoming. For reasons not immediately explainable more children in the upper age range than in the lower age range were assessed. The school counsellor

suggested that parents may be more willing to keep their children at home for an extra year if any difficulties are suspected. This artifact may also be explained, in part, by the fact that there were children already enrolled in Kindergarten, who were going to repeat. Some of these children were included in the screening and were in the upper age group. Determining reasons for the disproportionate age distribution is a fruitful area for further investigation.

Numbers

According to the counsellor at the school where the MAP was administered, approximately 280 children participated in the screening procedure. As previously described, almost all children entering kindergarten in September of 1989 were screened using the MAP. All children who were screened, and who subsequently remained in the school district throughout the 1989/90 and 1990/91 school years, were included. As noted, the MAP data were obtained by reviewing the records of every grade one student in the district. Demographic information was collected on each grade one child for whom a school registration form was found in the cumulative file, and MAP results, when present, were also recorded.

Data were collected on 345 students, 248 of whom had been administered the MAP. A count of children included on the last class lists indicated that 124 children were enrolled in the French Immersion program and 200 children were enrolled in the English program, during the 1989/1990 school year. This yielded a total enrollment of 324 in the E.C.S. program. Thus, MAP data were obtained for this study on approximately 77% of children enrolled in Kindergarten in the district during the 1989/1990 school year. Follow up data

were obtained on all grade one children enrolled in English language programs in the 1990/91 school year. 197 of those were children on whom the MAP scores had also been obtained (using the appropriate MAP form).

When MAP data were not available, several possible situations explained the lack. First, the grade one student may have enrolled in the school district after the screening had been accomplished (e.g., after the beginning of the kindergarten year, or any time before or during grade one. Second, the child may have been repeating grade one. In this case, any available pre-kindergarten screening would have been gathered using the old assessment instruments. Third, the child may have repeated kindergarten and not been included in the screening procedure, as the teachers already had a good understanding of his or her skills or developmental level. (Of the children whose MAP data were analysed 3.6 percent had repeated kindergarten, but had been included in the screening process. It was not known exactly how many children were repeating at the time of the initial screening but did not participate). Fourth, the child may have been invited to participate in the screening process and, for any number of reasons, the parent(s) declined having their child screened.

For the 23 percent of children enrolled in kindergarten for whom no MAP data were available, a number of additional explanations are possible. First, the child (for reasons suggested above) may not have been included in the screening procedures. Second, children enrolled in kindergarten in the school district may have moved from the district either before enrolling in grade one or during grade one before the data collection took place. Thus, the 77% of children for whom MAP data were available, represent children who were enrolled in kindergarten

and grade one in the same school district. Thus, generalizability is restricted to the proportion of students who remain in the school district over their first two years of school.

Of the 248 children for whom MAP data were available, complete data using the appropriate form for the child's age were collected on 197. 80 of these children had been administered the MAP IV and were therefore between the ages of four years three months and four years eight months. 117 children were appropriately screened using MAP V, and were therefore between the ages of four years nine months and five years two months.

Follow up achievement data were obtained on all children enrolled in English language grade one programs. Since children enrolled in French language programs were not administered any standardized achievement instruments, no follow up data were obtained. Of the children whose files were reviewed at the end of the grade one year, MAP data (from appropriate test forms) were available on 115 of the children who had Gates MacGinitie Reading Readiness test scores recorded, 109 of the children who had Canadian Test of Basic Skills test scores recorded, and 124 of the children who had Gates MacGinitie Reading Test scores recorded.

Instruments

Names

The following four instruments were administered to the sample previously described: The Miller Assessment for Preschoolers (MAP) developed by Miller (1985), the Gates MacGinitie Grade One Readiness Test (Gates Readiness) (Teachers College Press, 1966), the Canadian Test of Basic Skills (CTBS)

(Nelson Canada Limited, 1981), and the Gates MacGinitie Reading Test (Gates MacGinitie) (Nelson Canada Limited, 1980) (Level A, Form 1). These outcome measures represent the achievement measures used in the school district's assessment policy. Given that these are the measures actually used by the school board to monitor the progress of their students, they are the most appropriate outcome measures for this study.

Purpose

For the purpose of this study, the MAP scores will be considered an independent variable while the scores from the outcome measures will be considered dependent variables. However, since the analysis to be undertaken is correlational, this distinction is made solely to delineate the temporal relationship between the variables. That is, scores from the MAP precede in time the scores from the follow-up measures. As previously described the purposes of the outcome measures are to estimate reading readiness skills, and academic achievement.

Miller Assessment for Preschoolers.

The Miller Assessment for Preschoolers is a standardized test being used as a screening instrument in the hopes that the scores which it yields will be predictive of future academic achievement. Thus, to the extent that it predicts achievement there should be a correlation between scores obtained on the MAP prior to kindergarten, and scores on achievement measures administered in grade one.

Gates MacGinitie Reading Readiness Test.

This instrument is designed to measure the skills considered prerequisite to the reading process. Thus, it represents an intermediate level of assessment, between the early screening which examines a variety of developmental tasks, and the achievement measures which examine academic skills directly. The Gates Readiness is made up of seven subtests and yields one Readiness Standard Score, as well as stanine scores for each of the subtests. The subtests are labeled, and purportedly measure, Listening Comprehension, Auditory Discrimination, Visual Discrimination, Following Directions, Letter Recognition, Visual Motor Coordination, and Auditory Blending. Stanine scores from each subtest are weighted according to the formula provided by the publisher and an overall Readiness standard score is derived. This may also be reported as a percentile.

Canadian Test of Basic Skills.

The Canadian Test of Basic Skills is a widely used assessment instrument which provides estimates of achievement across a number of academic realms. On the form used with grade one children in this study, (Form 5, Level 6), the following subtests are included: (a) Listening, (b) Vocabulary, (c) Word Analysis, (d) Reading, (e) Language, and (f) Mathematics. This test also yields a Composite Score, reported as a grade equivalent, or as a percentile.

Gates-MacGinitie Reading Tests.

The Gates Reading Test measures reading achievement with respect to both Vocabulary and Comprehension. These two subtest scores yield a total standard score in the form of a T-score which may also be reported as a percentile. Thus,

its purpose is to provide a measure of reading achievement which considers both of these facets of reading and simultaneously provides one overall score.

Psychometric Properties

This section will describe the reliability and validity of each of the outcome measures. Each have unique strengths and weaknesses as outcome measures of achievement for this sample. These will be further discussed in the chapter five of the thesis where they impact the interpretation of the results of this study. Since the psychometric properties of the MAP have been discussed in detail in the chapter two, they are only briefly summarized in this section.

Miller Assessment for Preschoolers.

As noted in chapter two, the MAP manual reports an interrater reliability coefficient of .98 for the Total MAP score. Internal reliability estimates ranged from .79 to .82, which the author claimed to be reasonable "since the behavior sampled in the MAP is heterogeneous across motor and cognitive domains" (Miller, 1988, p. 106). Test-retest stability was reported in the form of stability ratios which indicate the percent agreement of overall classification across test administrations. 81% of children retested fell in the same classification category on both administrations. Content, criteria (concurrent and predictive), and construct validity of the MAP was described by Miller (1988). The reader may refer to chapter two for a complete discussion of these aspects of the MAP.

Gates MacGinitie Reading Readiness Test.

The teacher's manual for the Gates Readiness test states that the norms have been based on a nationwide (United States) sample, stratified on the basis of

geographic location, size, and socioeconomic level. No reliability or validity coefficients are provided in the Teacher's Manual. The technical manual reports subtest reliabilities but no total test reliability. Reliabilities for each subtest at each grade level were calculated using the KR-20 reliability coefficient. Median reliabilities ranged from a low of .63 for the auditory blending subtest, to a high of .87 for the visual discrimination subtest.

The technical manual also reports information concerning the validity of the Gates Readiness test as a predictor of Gates Reading Test scores. Readiness scores correlated at a level of only .26 with Reading Vocabulary scores, and .55 with Reading Comprehension.

Canadian Test of Basic Skills.

The Canadian Test of Basic Skills (Level 6, Form 5) was adapted from materials originally developed in the United States. In the early 1960's the Canadian project was begun at the University of Calgary. Earlier versions of Canadian forms were published in 1966, 1972, and 1974. Form 5, used in this study, was published in 1982 by Nelson Canada Ltd. Norms were based on a 1980 Canadian standardization. The normative group consisted of approximately 3200 pupils at each grade level. These pupils were sampled from schools in which English was the main language of instruction, and which was representative of all provinces and school sizes (King, 1982).

The Teacher's Manual for the CTBS (King, 1982) reports internal consistency reliability coefficients ranging from .75 to .92 for subtests across the different grade levels. All composite score internal reliabilities are reported to be above .90. The Manual for Administrators, Supervisors and Counsellors

provides more detailed descriptions of the K-R 20 internal consistency coefficients for form 6 (K-R 20 coefficients are reliability coefficients which reflect homogeneity of dichotomously scored items within a test (Sattler, 1988)). These range from .64 for the Vocabulary subtest to a high of .93 for the Composite score. The high internal consistency of the composite measure is evidence of adequate reliability of the CTBS as an outcome measure in the current study.

Evidence of validity is described in the Teacher's Manual as being based on many years of research in curriculum, measurement, and test interpretation. Furthermore, item selection was based on both empirical evidence of validity and judgements of a wide variety of curriculum experts. Evaluation by members of various ethnic groups were also included in item evaluation.

Gates MacGinitie Reading Tests.

The Teacher's manual for the Gates MacGinitie Reading Tests (Gates) (MacGinitie, 1980) describes the norming, reliability, and validity for the range of Gates MacGinitie Reading Tests. The Gates was normed on between 300 and 4500 children at each grade level. Students from all provinces and territories were included in the norm group. Each province or territory was proportionally represented with respect to population figures. In general, the sample within each province was stratified on the basis of city size, population, and type of school board (i.e., either public or separate).

Reliability for each form was determined using the Kuder-Richardson 20 formula, which provides a measure of internal consistency. These reliabilities were based on data from the Canadian Standardization sample. Separate

reliabilities are reported for both Vocabulary and Comprehension subtests. For level A the reliability of the Vocabulary subtest is reported to be .91 and for the Comprehension subtest, it is .92. Generally speaking reliabilities above .80 are considered adequate for group achievement tests, while reliabilities of at least .90 are expected for composite scores being used for placement purposes (Bracken, 1987).

Validity of the Gates is described in the Teacher's manual as it relates to curriculum content being assessed, and the representativeness of the items when compared to teacher expectations for reading skills. Users are encouraged to examine the tests and to determine if in fact the content reflects that emphasized and taught in their curriculum. No empirical research establishing the validity of the test is reported in this manual.

Scoring or Marking Procedures

Scoring and marking procedures for each of the instruments were reported to have followed the published standardized administration requirements. No deviations from these criteria were reported by the teachers administering the tests. Teachers recorded either total raw scores or normative scores such as grade equivalents or percentiles. For each outcome measure the normative form chosen was either recorded already or was derived from the information recorded by teachers and the tables provided in the manuals.

Administration Times and Conditions

Miller Assessment for Preschoolers. The Miller Assessment for Preschoolers was administered in ECS classrooms at the school which all kindergarten children in the district attended. The MAP was administered by

teachers trained in its use. The school counsellor and an occupational therapist received training in the use of the MAP directly from the instrument's developer, Lucy Jane Miller. Teachers were then trained in a workshop format by these individuals, using a videotape prepared by the tests developers. Practise tests were given under supervision and individual tutoring was provided.

Administration conditions fulfilled requirements stated in the test manual. Each child was brought to the school, usually by one or both parents, and was introduced to the examiner. A half hour play period was provided and then the MAP was administered by regular classroom kindergarten teachers. The actual administration time for the MAP varied somewhat, although for the most part the assessment took approximately thirty to forty minutes.

Gates MacGinitie Reading Readiness Test. The Gates MacGinitie Reading Readiness Test was administered by classroom teachers, in October of 1990, when most of the children had proceeded to grade one . According to the manual administration of this test takes approximately 120 minutes and can best be accomplished over 4 sessions spread over two different testing dates. As far as can be ascertained ng conditions were adequate. No teachers reported departures from standardized administrations or invalid testing conditions.

Canadian Test of Basic Skills. In February of 1991, when most of the children involved in the MAP screening were half way through their grade one year, the Canadian Test of Basic Skills was administered. According to the manual for this instrument, the entire administration takes approximately 205 minutes to administer. This time is comprised of 170 minutes of actual working time. Ideally this administration is completed in eight sessions spread over 5 testing

dates. Teachers were asked to make note of any administration irregularities or adverse testing conditions. None were reported.

Gates MacGinitie Reading Tests. The Gates MacGinitie Reading Tests were administered in May of 1991, toward the end of the grade one year. Thus, results of this instrument yield an outcome measure approximately two years later than the original screening. The manual for this test suggests that the administration is best accomplished over 2 sessions and takes approximately 90 minutes, of which 55 minutes is actual testing time. the remaining time is for distribution of materials, instructions, and practice items.

Data Collection and Recording

This section describes the methods used to aggregate the data for each child included in the study. For the most part it describes how the data on each student were accessed from various sources. The data were aggregated in two separate phases. First, the MAP data and school registration information were collected. Then, data from the follow up measures were aggregated. These two procedures are described separately in this section.

Miller Assessment for Preschoolers and school registration information.

The researcher began this procedure by identifying the information of interest contained in each student record. The district has a policy of retaining in records: (a) the MAP protocol, (b) a checklist completed by the parents, and (c) the school registration form. MAP data were recorded directly from the test protocol, including item scores and total scores.

It was decided that the most detailed demographic information could be obtained from the standard registration form. Thus, the researcher set up a data

coding format which allowed this information to be coded from the registration form, directly to a data coding sheet. The data were then transferred directly to a mainframe computer system by data entry personnel.

The demographic information obtained included: (a) date of birth, (b) parents' employment status (coded dichotomously as employed or not employed, for both mother and father), (c) parents' religious affiliation (coded as Roman Catholic, Christian - not Roman Catholic, Other, including any other stated affiliation, or None, including those who left the space blank or explicitly stated that they had no affiliation), (d) languages spoken at home (coded to allow for delineation of English, French, Other, and any combination of these), (e) number of siblings, coded to record gender and whether or not the siblings were older or younger, and (f) health and social history (coded to allow for each listed health history answer to be recorded as yes, no, or missing).

Outcome Measures. At the end of the school year, teachers had amalgamated the results of standardized assessments conducted over the year, or had placed these results in students cumulative files. Thus, the investigator went to each school and recorded birth date, gender, and student I.D. information (so results could be matched to MAP results), as well as the assessment results. Outcome measure results were then matched with MAP data; a unique I.D. number was assigned; the alphanumeric code was then dropped from the data set. Thus, the final data set contained no identifying information.

Data Analysis

Questions to be answered

1. What is the relationship between the scores on each of these measures? Are any associations between MAP scores and outcome measure scores statistically significant? Are the scores on the MAP predictive of scores on each of the standardized assessment outcome measures?
2. If the relationship between MAP scores and outcome measure scores is examined separately for groups with High, Medium, and Low performance on the MAP, are there differences between the correlations of the scores. That is, is the MAP a better predictor of achievement for the children in the lower third of the sample than it is for children in the middle, or upper thirds?

Statistical Techniques to be Employed.

In order to answer the questions regarding the associations between MAP scores and outcome measures, Pearson Product Moment Correlations were calculated. Separate correlations were calculated for each of the forms of the MAP (MAP IV and MAP V). Thus, initially six correlation coefficients were calculated (two for each of three outcome measures). These coefficients were considered statistically significant if the probability of finding such a correlation by chance, was less than .05 (i.e., $\alpha \leq .05$). The amount of variance in outcome measure scores accounted for by variance on the MAP was estimated by calculating the square of the correlation coefficient.

In order to answer question number two above, the data were examined in order to determine how best to divide the MAP scores into High, Medium, and Low groups. Since the goal was to identify groups which might exhibit differing

levels of association to the outcome measure, a decision was made to use a proportion of children scoring below the mean and an approximately equivalent proportion of children scoring above the mean, whilst simultaneously producing a middle group. This procedure was predicted to create groups where differences in the association amongst screening and outcome measure scores would be identifiable. That is, if the associations are stronger in the extremes of the group, or in the middle, any such variations would be identifiable. Raw score cutoffs for each Level (IV and V) of the MAP were determined which would identify approximately the lower and upper quarter to third of children on whom data were available for each outcome measure. Using the same cutoff score each time allows for consistency in interpretation of each analysis. However, this yielded somewhat different subsample sizes and proportions for each of the analyses. As well, due to differences in the distributions of scores between the MAP IV and the MAP V, different cutoff scores were used for the two instruments.

Since the data were in the form of raw scores on the MAP and in the form of standard scores on the outcome measures, all instruments yielded interval level scores. This allowed for Pearson Product Moment correlations to be computed and compared. Correlation coefficients for the three groups were calculated and when appropriate compared using the Fisher' z test. This technique allows the researcher to test for statistically significant differences between two correlation coefficients. Results were considered significant if the size of the difference was unlikely to have occurred by chance alone. Differences which occur less than .05 of the time by chance alone were considered significant (i.e., $\alpha \leq .05$).

IV. Results

Introduction

This chapter will describe the results of the empirical data analysis. Results for each outcome measure instrument will be examined individually, in the order in which the instruments were administered throughout the grade one year. This corresponds to the prediction interval from shortest to longest. Thus, the analyses relating to the Gates MacGinitie Reading Readiness Test are presented first, followed by those for the Canadian Test of Basic Skills, and finally by those related to the Gates MacGinitie Reading Tests. Within each section the results are presented separately for the two age groups. That is, results are presented first for the younger children, assessed using Form IV of the MAP, followed by those for the older children, assessed using Form V of the MAP.

Table 4.1 presents the results of all the analyses to be described in the balance of this chapter. Correlations between each MAP form and each outcome measure are presented with and without outliers. These are discussed in the order described in the following sections of chapter four.

Table 4.1

Correlations between MAP Scores and Outcome Measure Scores

MAP IV		MAP V	
Readiness		Readiness	
Total Sample		Total Sample	
with outlier	.32 *	with outlier	.40 *
without outlier	.20	without outlier	.32 *
Low Group		Low Group	
with outlier	.58 *	with outlier	.48 *
without outlier	.40	without outlier	.34
Medium Group		Medium Group	
	-.11		.22
High Group		High Group	
	.47 *		.13
CTBS		CTBS	
Total Sample		Total Sample	
with outlier	.43 *	with outlier	N/A
without outlier	.29 *	without outlier	.39 *
Low Group		Low Group	
with outlier	.35	with outlier	N/A
without outlier	-.09	without outlier	.38 *
Medium Group		Medium Group	
	.21		.33
High Group		High Group	
	.20		.33
Gates Reading		Gates Reading	
Total Sample		Total Sample	
with outlier	.13	with outlier	.21 *
without outlier	.09	without outlier	.22 *
Low Group		Low Group	
with outlier	.44	with outlier	.14
without outlier	-.08	without outlier	.05
Medium Group		Medium Group	
	.11		-.07
High Group		High Group	
	-.09		.27

Note. All reported correlations are Pearson Product Moment coefficients
 * denotes significant at $\alpha=.05$

Gates MacGinitie Reading Readiness Test

Form IV.

Readiness Standard Scores were available for 41 of the 80 children on whom MAP Form IV data were appropriately recorded. Since approximately 40% of the original group administered the MAP went into French Immersion grade one programs, it is likely that this figure represents approximately 85% of the children for whom follow up data were expected. A Pearson Product Moment Correlation coefficient of 0.32 was calculated. This describes a weak but statistically significant ($\alpha=.02$) relationship between scores on the MAP and the overall Readiness Standard Scores (RSS). It also indicates that ten percent of the variance in RSS is explained by variation in the scores on the MAP ($0.32^2=0.10$).

Upon examination of the scatterplot of the scores, the size of this correlation coefficient appeared to be substantially affected by the presence of one outlier. This subject's score on the MAP IV was more than one standard deviation lower than the next lowest score in the sample. Thus, since it was possible that the score represented any number of possible measurement errors, or the subject was not actually part of the population of interest, the correlation was recalculated in the absence of this data point. Removing the outlier yielded a Pearson Product Moment correlation of .20. which was not statistically significant and may have occurred by chance in the absence of any linear relationship between the scores.

In order to determine if there was a stronger relationship between lower scores on the MAP and the Readiness measure, than between higher scores on the

MAP and this measure, the group was divided into three groups. The groups were divided according to whether the children had obtained raw scores on the MAP IV of 72 or below, from 73 and 77 (inclusive), or 78 and above. These cutoff scores were chosen in order that the same cutoffs could be used for each analysis while yielding approximately one quarter to one third of the sample in the lower and upper groups and one half to one third of the subjects in the middle group.

Division of the sample in this manner yielded a group of twelve students with relatively low scores on the MAP, and a group of fifteen children with relatively high scores. Twenty-four children having middle range scores on the MAP formed the middle group.

Pearson Product Moment correlations were calculated separately for each of these groups. For the lower group the correlation between MAP scores and Gates Readiness scores was 0.58. This is statistically significant at the $\alpha=.025$ level. However, as noted above, the size of this correlation coefficient was substantially inflated as the result of one outlying data point. When this subject's scores were excluded from the analysis the resultant correlation coefficient was $r=.40$. This coefficient did not reach statistical significance in part due to the small size of the sample (now reduced from twelve to eleven). The correlation coefficient between MAP scores and the Gates Readiness for the upper group was 0.47 which is statistically significant at the $\alpha=.04$ level. Thus, the coefficients for both the upper and lower groups were larger than that for the group as a whole. The correlation between scores on the MAP IV and the RSS for the group of students who had middle range scores on the MAP, was $-.11$ which did not differ significantly from 0. Thus, it appears that

while a statistically significant relationship was demonstrated between the scores on the measures for the upper group, the scores for the middle and lower groups were not demonstrably different from 0.

Form V.

Readiness Standard Scores were available for 63 of the 117 children on whom MAP Form V data were appropriately recorded. Since approximately 40% of the original group who were administered the MAP went into French Immersion grade one programs, it is likely that this figure represents approximately 90% of the children for whom follow up data were expected. A Pearson Product Moment Correlation coefficient of 0.40 was calculated. This describes a weak to moderate, and statistically significant ($\alpha=.001$) relationship between scores on the MAP and the overall RSS. It also indicates that sixteen percent of the variance in RSS is explained by variation in the scores on the MAP ($0.40^2=.16$).

As for the MAP IV, this correlation coefficient was noted to be substantially affected by the scores of a single low outlier. In this case the MAP score for the outlier fell more than four standard deviations below the next lowest score. Removing this outlier yielded a correlation coefficient of .32 which suggests a weak, yet statistically significant ($\alpha=.005$) relationship between scores on the two measures. In contrast to the effect for the MAP IV results when removing the outlier, the scores on the MAP V and RSS yield a correlation coefficient which remains significant in the absence of the outlier.

In order to determine if there was a stronger relationship between lower scores on the MAP and the Readiness measure, than between higher scores on the

MAP and this measure, the group was divided according to whether the children had obtained raw scores on the MAP of 73 or below, from 74 to 76 inclusive, or 77 and above. Again, these cutoff scores were chosen in order that the same cutoffs could be used for each MAP V analysis while yielding approximately one quarter to one third of the sample in each group.

Division of the sample in this manner yielded a group of nineteen students with relatively low scores on the MAP, and a group of twenty-one children with relatively high scores. Twenty-three children with middle range scores on the MAP comprised the middle group.

Pearson Product Moment correlations were calculated separately for each of these groups. For the lower group the correlation between MAP scores and RSS was 0.48. This is statistically significant at the $\alpha=.02$ level. However, when the scores of the outlier were removed from the calculation, the coefficient dropped to .34, which did not reach significance at the .05 level. Thus, no linear relationship was conclusively demonstrated. For the middle group (N=23) the correlation was .22. This is not a statistically significant finding in a sample of this size, and may have occurred due to chance. The correlation coefficient between MAP scores and the Gates Readiness for the upper group was 0.13. This correlation coefficient is not statistically significant.

When the outlier is included in the analysis, there appears to be a stronger relationship between the measures for those in the lower group on the MAP than there is for the higher group. This would suggest that the relationship between scores on these measures is not linear in form. However, when the

outlier is removed, only the weak correlation of scores for the whole group is statistically significant.

Canadian Test of Basic Skills

Form IV.

Grade equivalent scores, which have normal curve characteristics, were available for 43 of the 80 children on whom MAP Form IV data were appropriately recorded. Since approximately 40% of the original group administered the MAP went into French Immersion grade one programs, it is likely that this figure represents approximately 90% of the children for whom follow up data were expected. A Pearson Product Moment Correlation coefficient of 0.43 was calculated. This describes a weak to moderate, and statistically significant ($\alpha=.002$), relationship between scores on the MAP and the overall grade equivalent scores on the CTBS. It also indicates that eighteen percent of the variance in grade equivalent scores is explained by variation in the scores on the MAP ($0.43^2=0.18$).

Upon examination of the scatterplot of the scores, it appeared that the scores of one outlier were substantially affecting the magnitude of the correlation coefficient. When scores of this outlier were removed from the analysis, the correlation coefficient dropped to .29 remained still significant at the .05 level ($p \leq .003$). Thus a weak relationship between scores on these measures was demonstrated. However, in the absence of the outlier, the percentage of variance of CTBS scores accounted for on the basis of MAP IV scores, was only 8.4 ($.29^2=.841$).

In order to determine if there was a stronger relationship between lower scores on the MAP and the CTBS, than between higher scores on the MAP and the CTBS, the group was divided according to the same rules as used for the analysis described above for the MAP IV and Gates Reading Readiness Standard Scores. These cutoff scores were chosen in order that the same cutoffs could be used for each analysis while yielding approximately one quarter to one third of the sample in each group.

Division of the sample in this manner yielded a group of thirteen students with relatively low scores on the MAP, and a group of thirteen children with relatively high scores. Twenty-five children with middle range scores on the MAP made up the middle group.

Pearson Product Moment correlations were calculated separately for each of these groups. For the lower group the correlation between MAP scores and CTBS scores was 0.35 which does not reach statistical significance in a sample of this size. When the results of the outlier were removed from the analysis, the correlation dropped to -.09. This correlation was not significantly different from 0. The correlation coefficient between MAP scores and the CTBS for the upper group was 0.20. This coefficient is not statistically significant. Thus, only the weak significant relationship for the whole group was demonstrated for children administered form IV of the MAP and the CTBS.

Form V.

Grade equivalent scores from the CTBS were available for 67 of the 117 children on whom MAP Form V data were appropriately recorded. Since approximately 40% of the original group who were administered the MAP went

into French Immersion grade one programs, it is likely that this figure represents approximately 95% of the children for whom follow up data were expected. A Pearson Product Moment Correlation coefficient of 0.39 was calculated to describe the relationship between these two measures. This describes a weak to moderate and statistically significant ($\alpha=.001$) relationship between scores on the MAP and the overall grade equivalent scores on the CTBS. It also indicates that fifteen percent of the variance in RSS is explained by variation in the scores on the MAP ($0.39^2=.15$).

In order to determine if there was a stronger relationship between lower scores on the MAP and the CTBS composite score, than between higher scores on the MAP and this measure, the group was divided according to whether the children had obtained raw scores of 73 or below, from 74 to 77 inclusive, or 78 and above. These cutoff scores were chosen in order that the same cutoffs could be used for each analysis while yielding approximately one quarter to one third of the sample in each group.

Division of the sample in this manner yielded a group of twenty-one students with relatively low scores on the MAP, and a group of twenty-five children with relatively high scores. Twenty-three children with middle range scores on the MAP comprised the middle group.

Pearson Product Moment correlations were calculated separately for each of these groups. For the lower group the correlation between MAP scores and CTBS grade equivalent scores was 0.38. This is statistically significant at the $\alpha=.05$ level. The correlation coefficient between MAP scores and the CTBS grade equivalent score for both the middle and the upper groups were 0.33

which approached, but did not reach, statistical significance. These correlations are not statistically significantly different (Fisher's $Z=.54$, $p \leq .29$).

Gates MacGinitie Reading Tests

Form IV.

T scores for the total Gates MacGinitie Reading Test, which have normal curve characteristics, were available for 46 of the 80 children on whom MAP Form IV data were appropriately recorded. Since approximately 40% of the original group who were administered the MAP went into French Immersion grade one programs, it is likely that this figure represents approximately 96% of the children for whom follow up data were expected. A Pearson Product Moment Correlation coefficient of 0.13 was calculated. Since this coefficient was not statistically significant at the $\alpha=.05$ level, no conclusion can be drawn that there is in fact a relationship between MAP Form IV scores and reading achievement (as measured by the Gates MacGinitie Reading Tests) at the end of grade one. When the effects of one outlier were removed from the analysis the magnitude of the correlation coefficient dropped even further and was $-.09$ which was not significantly different from 0.

In order to determine if there was a stronger relationship between lower scores on the MAP and the Readiness measure, than between higher scores on the MAP and this measure, the group was divided according to the method described above for the parallel analyses using the Gates Readiness scores and the CTBS grade equivalent scores.

Division of the sample in this manner yielded a group of twelve students with relatively low scores on the MAP, and a group of sixteen children with

relatively high scores. Twenty-two children with middle range scores on the MAP comprised the middle group.

Pearson Product Moment correlations were calculated separately for each of these groups. For the lower group the correlation between MAP scores and Gates Readiness scores was 0.44. This coefficient closely approached, but did not reach, statistical significance due to the small sample. However, when the scores of one outlier were removed from the analysis, the magnitude of the correlation dropped significantly yielding a coefficient of -.08. Again, this coefficient is not statistically significant. The correlation coefficient for MAP and Gates Reading scores, for the upper group was -0.09. The correlation between scores on the MAP IV and T scores on the Gates Readiness Tests for the middle group was .11. Again, this coefficient is not significantly different from 0. These results suggest that, for those children administered MAP IV, there is no linear relationship between scores on that instrument and Gates Reading Test scores.

Form V.

Total T scores on the Gates Reading Test were available for 72 of the 117 children on whom MAP Form V data were appropriately recorded. Since approximately 40% of the original group who were administered the MAP went into French Immersion grade one programs, it is likely that this figure represents approximately 98% of the children for whom follow up data were expected. A Pearson Product Moment Correlation coefficient of 0.21 was calculated. This describes a weak but statistically significant ($\alpha=.04$) relationship between scores on the MAP and the overall T scores on the Gates. A correlation of this magnitude indicates that only four percent of the variance in T

scores is explained by variation in the scores on the MAP ($0.21^2=0.04$). When the scores of one outlier were removed from the analysis, the resultant coefficient was .22 which remained statistically significant at the .05 level ($p \leq .03$). This coefficient explains only 4.8 percent of the variance in T scores on the basis of MAP V scores ($.22^2=.048$).

In order to determine if there was a stronger relationship between lower scores on the MAP and the Readiness measure, than between higher scores on the MAP and this measure, the group was divided in the manner described for the previous parallel analyses.

Division of the sample in this manner yielded a group of twenty-two students with relatively low scores on the MAP, and a group of twenty-five children with relatively high scores. Twenty-five children with middle range scores on the MAP comprised the middle group.

Pearson Product Moment correlations were calculated separately for each of these groups. For the lower group the correlation between MAP scores and Gates T scores was 0.14. For the middle group the correlation was -.07. The correlation coefficient between MAP scores and the Gates T scores for the upper group was 0.27. None of these coefficient reached statistical significance. Thus no linear relationship between scores on the MAP V and Gates Reading Tests was demonstrated.

Summary

In summary, when the effects of outliers are removed, some weak correlations were found between the MAP scores and the outcome measures. Significant correlations of MAP IV scores with RSS, as well as with CTBS grade

equivalent scores were demonstrated. MAP IV scores did not correlate significantly with results of the Gates Reading Test administered at the end of the grade one year. Significant correlations between MAP V scores and each of the outcome measures were demonstrated, however these ranged only from .22 to .39, explaining little of the variance in outcome measure scores on the basis of the MAP results.

V- Discussion

Introduction

This chapter will discuss the results presented in chapter four as they relate to the research questions. This discussion includes comment on the limitations and implications of these results. The description of limitations refers to both the limitations of the design, the sample, and the impact of the results in terms of overall evaluation of the MAP as a kindergarten screening measure. A summary of the pertinent results and discussion issues is then followed by a statement of conclusions and recommendations made based on this study. Directions for further research are included where appropriate.

Discussion

Discussion of Results

The results described in chapter four suggest that MAP scores are weakly correlated with most outcome measures. Although some correlations reached moderate absolute values, when the effect of distinct outliers were removed these relationships were reduced to weak and often insignificant levels.

MAP IV

The MAP IV results are pertinent to the subsample of the population representing children between four years three months of age and four years eight months of age.

Gates Reading Readiness Test.

Overall, the MAP IV correlates with this outcome measure only when the score of an outlier is included in the analysis. In the absence of this outlier, no significant correlation was obtained for the group as a whole. In addition, the

moderate correlation for the low group also is reduced to insignificance when the outlier is removed. Interestingly, the relationship between MAP IV scores and Readiness Standard Scores is significant for the group who obtained high scores on the MAP. This is contrary to expectations, since the MAP is designed to be less discriminatory at the upper level. With respect to the usefulness of this correlation for the school district's purposes, it is unlikely that this provides information of significant value. The MAP does not purport to differentiate among children of average or above average ability, and it has not been a goal of the district's screening program to do so. Neither is the moderate correlation ($r = .48$) of sufficient magnitude to interpret good predictive validity for this group.

Canadian Test of Basic Skills.

Scores on the MAP IV are weakly correlated with scores on the CTBS ($r = .29$) even when the inflating effects of a single outlier are removed. However, despite the statistical significance of this relationship the strength of the correlation is not high enough to justify prediction for any individual child, regardless of the level at which he or she scored on the MAP IV. That is, when the MAP scores are divided into three groups, the correlations with CTBS scores are weak and not statistically significant. Despite the overall significant relationships the results of this analysis do not lend support to the MAP IV's validity as a screening measure which predicts overall academic achievement in grade one. Although the correlations are not statistically significantly different for the readiness measure and the CTBS (Low group, without outlier), the CTBS correlation is lower. This is consistent with the expectation that correlations between

measures would decline as the prediction interval increases. In this case, the CTBS was administered approximately five months later than the Readiness measure. As well, the CTBS measures academic skills directly, and is thus conceptually further away from the prerequisite skills measured by the MAP than are the readiness skills measured by the Gates Reading Readiness Test.

Gates Reading Test.

Consistently with expectations, as the outcome measure tasks become further removed from the screening tasks, and the prediction interval lengthens, the correlations become weaker and less likely to reach significance. In the case of the MAP IV scores and Gates Reading Test scores (administered two years after the initial screening) the correlation for the overall group is near to zero ($r = .13$). No correlations for any subgroup (with or without outliers) is statistically significant. Thus it appears that the MAP IV is not useful with this population for predicting reading ability two years after the initial screening. These results are consistent with the hypothesis that most learning disabilities (and presumably poor reading achievement) are language based and the MAP IV is ineffective in identifying such underlying problems, given its dearth of language related items.

The degree of association tends to increase when only the children with relatively low scores on the MAP are considered. The trend is for the association between MAP scores and measures of academic achievement to decrease as the length of time between assessments increases. This is expected in terms of development and measurement theory, given that historical and maturational factors intervene. However, as different instruments were used at the different

prediction intervals it is unknown in this case whether the different levels of association are related to chance, the time interval, the reliability of the measures, the skills being measured, an interaction of these factors, or extraneous factors to any of the above. However, these results do suggest that when Reading skills are considered (as measured by the Gates Reading Tests), the predictive ability of the MAP over a two year time frame is negligible.

MAP V

The MAP V results are pertinent to the subsample of the population representing children between four years nine months of age and five years three months of age.

Gates Reading Readiness Test.

As predicted the MAP V produced a stronger correlation with Readiness scores than did the MAP IV. This suggests that the older children's scores may be more able to predict readiness scores than those of younger children. Nevertheless, in the absence of outliers, the correlation between readiness scores and MAP scores is weak ($r = .32$) though statistically significant. The strength of this correlation in no way supports the use of this predictor upon which to make diagnostic, placement, or programming decisions.

The correlations are only significant when considering the entire group (that is, the group of children four years nine months to five year three months of age at the time of the screening). When subsamples based on MAP V scores were examined the correlations for none of the groups reached significance in the absence of outliers. This was due to the reduced sample size for the low group, since the actual magnitude of the correlation was similar to that for the group as

a whole. Thus, as a predictor of reading readiness skills, while the MAP V does produce a positive and significant relationship overall, the strength of this relationship is insubstantial for practical, decision making purposes.

Canadian Test of Basic Skills.

MAP V scores correlate significantly with CTBS scores regardless of the fact that no outlier artificially inflates the correlation in this case. The correlation is also statistically significant for the relationship between scores for those who scored in the low group on the MAP and not for those who scored in the middle or high groups. Since the difference in the magnitude of the correlations is very small, the lack of significance obtained in the middle and high groups relates to the sample sizes and the fact that the low group correlation only just reached significance. These results are as predicted however, with respect to overall trends. That is the MAP V scores are significantly related for the group amongst whom the MAP is expected to differentiate most strongly.

Gates Reading Tests.

In contrast to the results for the MAP IV, the MAP V does correlate significantly with the Reading Test results. However, this significant correlation is interpretable in part, as a function of the somewhat larger sample size for this group. None of the smaller subsample correlations (based on magnitude of MAP scores) reached statistical significance. Regardless of the statistical significance, the weak correlations imply no practical predictive value. Again, the size of the correlations obtained does not provide evidence for the usefulness of the MAP as a measure upon which to base decisions.

Hypotheses Regarding Results

The differences between the skills measured on the MAP and on the outcome measures appear to be substantial. The three outcome measures each have strengths and weaknesses as reference points in determining the validity of the MAP as a screening instrument. The readiness measure is a useful outcome measure in that it provides a measure of skills which can be conceived of as a bridge between screening measure skills (e.g. neuromaturational items), and actual academic skills (e.g. reading or arithmetic skills). As such, the readiness measure reflects the child's skills on items for which the MAP skills might be prerequisites.

Achievement items may rely on the acquisition of both MAP skills and readiness skills as measured by these instruments. That is, if the theoretical basis of the readiness measure is accurate, achievement skills will rely on abilities reflected by the readiness measure. The weakness of the readiness measure is that it does not provide of itself a measure of the end construct of interest, i.e. to the extent that the goal of the screening is to predict achievement, the readiness measure does not provide a meaningful estimate of the MAP's ability to do so.

The Canadian Test of Basic Skills does provide such a measure of achievement. The composite score is a global measure of achievement across the major curricular domains of serious consideration in the early elementary years. Teachers can be expected to be particularly interested in the achievement of their students in the area of basic skills, such as reading, language, and arithmetic. The CTBS's strength therefore is in its ability to provide a correlation coefficient

which will reflect the screening instrument's relationship to overall academic achievement. Its weakness as an outcome measure is related to the weaknesses of group achievement tests to reliably and validly reflect actual academic achievement.

The Gates Reading Test provides a useful outcome measure in that it yields an indication of the extent to which the MAP predicts scores for a specific and vital aspect of academic achievement. Reading achievement is likely to be one of the primary areas of interest of elementary teachers, given the relationship of reading skills to all areas of academic achievement. However, the correlation between reading and screening results is likely attenuated by the distinctly disparate natures of the tasks measured on the two instruments. The MAP results are likely to be least correlated with the most specific measure of achievement which is represented by the reading tests.

The time factor between assessments would be expected to relate to the strength of the correlation coefficients. That is, the correlations between screening measures and outcome measures is expected to diminish as the time interval increases, given the increased importance of historical events and developmental discontinuity. Thus, it would be expected that the screening measure would be more highly correlated with the Readiness measure which represents the shortest inter-test interval, followed by the CTBS which represents the intermediate interval, and would be lowest with the reading achievement measure which represents the longest prediction interval (almost exactly two years).

The MAP measures skills which are indirectly related to academic achievement. Many items are based on fine and gross motor skills, and load on perceptual and visual-motor factors. Relatively few items (4 of 27) measure language based skills. In contrast, the outcome measures are examining skills and achievement with an apparently heavier load on language based skills. For example, the Readiness measure examines four language related prerequisite skills to the reading process, including listening comprehension, following directions, auditory discrimination, and auditory blending. In contrast, visual and perceptual skills are measured on two subtests examining letter recognition, and visual discrimination. Further research might determine if in fact scores on the MAP are more strongly related to these two visually oriented subtests, than to the balance of the Readiness tests described above.

Limitations

The limitations of this study are many and varied, although they do not jeopardize the overall usefulness of the study. First, the study involves post-hoc analysis of data collected for other purposes. Since this is the case, the *ideal* instruments and measures with which to validate the MAP have not necessarily been chosen. For example, it is possible that in a future study more reliable and valid measures, such as individual achievement tests, could be used. However, since the instruments used do meet minimum criteria for group administered tests, it appears likely that individualized tests of achievement would be necessary in order to improve substantially on the psychometric characteristics of the instruments used.

Also an artifact of the post-hoc design is the fact that complete descriptive information for the sample was not available. Therefore, information with respect to subject characteristics of interest, such as parents occupations and socio-economic status, custody arrangements, and ethnic background, was unavailable. Nevertheless, as the subjects included in the results represent a majority of students who enrolled in the district's English language programs in both kindergarten and grade one it is reasonable to suggest that the results are generalizable to that population in future administrations of the MAP.

The time frame of this study prohibits the examination of follow-up data collected more than two years after the original administration of the MAP. Thus, the predictive validity being tested in this study represents a relatively short term prediction interval. However, since these results suggest that even this interval leads to highly questionable predictive validity it is likely that longer intervals would lead to even lower correlation coefficients. In fact this trend appears in the data reported above. Nevertheless studies of predictive validity of any screening or diagnostic measure are needed on a longer term basis for complete justification of their use.

There is utility to deriving information with respect to predictive validity over the time period proposed. Since predictive validity usually declines over time, given the uneven nature of development and the imposing forces of environmental and situational factors, relatively short term studies such as this one are also needed. While these results may be problematic with respect to the fact that not all academic deficits can be identified at an early stage in grade one, the combination of results from both long and short term studies is necessary.

As this is an initial level inquiry into the predictive validity of the MAP, it is limited in the extent of the statistical analysis of the data which has been completed to date. The evaluation of the correlation coefficients presented in no way represents the full extent of the analysis necessary to determine the MAP's usefulness for a variety of purposes. Hit-rate analyses and carefully chosen and followed referral mechanisms based on MAP results, are necessary in order to evaluate its use as a screening instrument in the school district.

Implications

The primary implication of the results of this study relate directly to screening procedures and the policy in the school district in question. Using a sample which represents children enrolled in both kindergarten and grade one, the present study suggests that overall correlations between MAP results and readiness and achievement results are generally weak. Thus, while the screening procedure is designed to identify children who are at risk for low academic achievement, the results of this data analysis suggest that for this sample the MAP scores are not highly correlated with academic achievement in grade one. That is, children with low scores on the MAP do not necessarily obtain low outcome measure scores, and children who score well on the MAP do not necessarily score well on outcome measures.

Numerous explanations for the patterns of correlations observed are possible. First, the reliability of each measure directly affects the maximum correlation coefficient which it is possible to obtain. Since the MAP has reliability figures which are difficult to interpret, the choice to examine the internal consistency reliability seems most practical in examining this factor.

The MAP reports a split-half reliability of .79 which its author claims to be reasonable given the heterogeneous nature of the items. However, this reliability does not meet minimum criteria as stated by measurement experts. As well, an instrument with such a reliability would be expected to correlate with itself only at a level of 62, at best ($.79^2 = .62$). Thus, the lack of reliability of the MAP itself mitigates against its usefulness as a predictor of academic achievement. Error in the scores obtained from the outcome measures also contributes to attenuated correlations even when a linear relationship between true scores actually exists (Nunnally, 1970).

The fact that the MAP results were used in part to determine classroom placements for the kindergarten children, suggests that a treatment effect may be attenuating the correlation of screening and achievement results. However, while this explanation of low correlations should not be ruled out entirely, it is true that the classroom placements were only affected by class size, and no specific individual program plans were devised on the basis of MAP results. Therefore no specific differential treatment was implemented, and children not placed in small classes on the basis of MAP results also benefitted from the smaller class sizes. (It should be kept in mind however, that the effect may have been differential between those selected on the basis of MAP results, and those not. That is, children selected on the basis of MAP results may in fact have derived proportionally greater benefit from being enrolled in smaller class sizes, thus attenuating the correlations.)

The correlations may also be weak because there is actually little relationship between skills and abilities measured by the MAP and those

measured by the outcome instruments. Developmental considerations, as well as the theoretical importance of the construct being measured are both important determinants in this argument.

First, preschool screening in itself rests on the sometimes dubious assumption that skills develop evenly over time and that children who exhibit weak skills prior to kindergarten are at significantly greater risk for exhibiting weak skills or abilities later in their school years. This assumption has been the cause of great controversy, and much has been written regarding the usefulness of any screening procedure, given the reality of uneven developmental patterns of growth in young children (e.g. Lichtenstein and Ireton, 1984).

Meisels (1987) notes that a maturationist theory of development is the basis of such screening and readiness tests such as the Gesell School Readiness Screening Test. Such a theory suggests that time is the crucial variable predicting changes in observable behaviour. Meisels suggests that this is an outmoded theory of development which fails to take into account the numerous other factors which affect behaviour. Thus, screening and readiness tests which presume that maturation is the sole factor predicting academic achievement, fail to allow for the numerous other variables (such as experience and socioeconomic variables) which will impact a child's school learning abilities. Meisels advocates the use of screening and readiness measures for the purpose of leading to comprehensive assessment and to make "better and more appropriate services available to the largest number of children" (1987, p.71).

Since the lack of correlations found in this study imply the inadequacy of the MAP in predicting academic achievement, the implications, combined with such a

view as Meisels, are that great care needs to be taken in the district to ensure that results of the MAP lead to such a goal for the individual child. Thus a clear policy of screening and use of such results needs to be developed.

Second, screening using the MAP assumes that the constructs measured by the MAP are directly related to the acquisition of academic skills or abilities. As pointed out by Padget (1985) and as recognized by Miller (1988) the MAP assesses predominantly visual perceptual and motor skills, while most learning disabilities are related to language based disorders. Of the 27 items included on the MAP only 4 are included on the Verbal index. Thus, there is theoretical explanation for relatively low correlations between MAP results and measures of academic skills. Satz and Fletcher point out that early evidence suggested that visual-spatial constructs were better predictors of later reading ability than language related measures. However, these authors proceed to describe a number of further studies which found that a number of language related measures, not included in the early research, were predictive of achievement generally. Further, they cite Jorm, Share, MacLean, and Matthews (1986) who found that specific measures of phonologically related skills were predictive of reading disabilities. Overall, their call for further research into the best predictors of specific learning disabilities, remains unanswered.

The evidence suggests that the results of the MAP as collected with a majority of kindergarten children enrolled in the school district, are not highly related to further academic achievement. Thus, these results warrant the careful reconsideration of the deployment of resources currently being used to implement the screening procedure as it now stands. Numerous authors have

reviewed research and policy with respect to preschool screening and have developed guidelines for implementing screening procedures. The results of the current investigation indicate that careful consideration of the goals of screening need to be undertaken, as in and of itself the MAP is not predicting with accuracy the future academic achievement of students undergoing the screening procedure.

Some of the factors to be considered in screening policy are described below. However, it should be noted that in this context it is not possible to fully review the plethora of published material which describes issues in screening and preschool assessment. Instead, the main conclusions of a limited number of review articles which highlight main issues and provide criteria for preschool assessment procedures have been amalgamated as succinctly as possible.

Preschool Screening Models and Practice

Readiness, Screening, and Diagnosis

Screening has been defined as the "identification of unrecognized disease or defect by the application of tests, examinations, or other procedures which can be applied rapidly and are relatively simple and reliable with few false positive or false negative results." (Robertson and Cockburn, 1974 as cited in Drillien, Pickering, and Drummond, 1988 p.294). Others have referred to screening as involving the evaluation of large groups of children with brief, low cost procedures to identify children at risk for academic problems (Paget and Nagle, 1986). Meisels (1987) differentiates this process from that of readiness testing which is asserted to be assessments of curriculum related skills that a child has acquired and that are considered prerequisite skills for further academic development. Further, it is vital to differentiate these from diagnostic

assessment or evaluation, which Paget and Nagle (1986) and other authors (see for instance Taylor, Willits, and Lieberman, 1990 as well as general texts such as Lichtenstein and Ireton, 1984, and Salvia and Ysseldyke, 1985). Any preschool assessment policy or program should be aware of these distinctions and screening procedures should use screening tests, while readiness measures should be used to investigate readiness for the purpose of program planning and more comprehensive assessments are necessary for diagnostic evaluation.

Models of Preschool Screening

Although varying models for preschool screening have been proposed, many share core characteristics. Of the models or review articles which concisely address many screening issues (NAEYC, 1988; Lehr, Ysseldyke, and Thurlow, 1987; Paget and Nagle, 1986; Taylor, Willits, and Lieberman, 1990; Wilson and Reichmuth, 1985; Stevenson, 1984; Satz and Fletcher, 1988), many common themes are identified.

Multiple Information Sources

The National Association for Education of Young Children (1987) provides guidelines for appropriate practice and include the primary recommendation that:

Decisions that have a major impact on children,, such as enrollment, retention, or placement, are not made on the basis of a single developmental assessment or screening device but consider other relevant information, particularly observation by teachers and parents. Developmental assessment of children's progress and achievements is used to adapt curriculum to match

the developmental needs of children, to communicate with the child's family, and to evaluate the program's effectiveness. (p. 13)

Similar requirements exist for multiple sources of information in situations where there will be substantial impact on the child in Alberta's School Act.

Observation in various situations, particularly the home, is advocated by author's such as Taylor, Willits, and Lieberman(1990) and Paget and Nagle (1986). Overall, there appears to be a consensus that even the best psychometric estimate of a child's developmental status cannot stand alone in determining the need for delayed entrance, special programming, or placement.

Adequate Reliability and Validity of Instruments

Above all there is consensus that reliability and validity of the measures to be used in a screening program must be firmly established. This requires careful examination and *critical* review of the psychometric information provided by the test publisher, as well as perusal of independent research and reviews, as well as carefully planned and implemented local research which will establish the validity of the instrument used for the specific purpose in the specific region in which it is used.

Criteria for evaluating preschool assessment measures have been put forth (Lehr, Ysseldyke, and Thurlow, 1987). These indicate that for screening purposes, measures must meet reliability criteria of at least .80, and that the type of reliability should be clearly identified in the test publication material. When any placement decision is being made, reliability must reach .90. Requirements for norms, and validity are similarly presented by these authors. Evidence of construct, content, and predictive validity should be presented in a

manner which allows for critical inspection of the claims to validity. Generally, the criteria proposed by these authors conform to the more general Standards for Educational and Psychological Tests, as mandated by the American Psychological Association (1985).

Referral for Diagnostic Assessment

Since screening measures are not intended to diagnose specific disabilities, nor will they result in perfect hit-rates (identification of all at risk children and no identification of children not at-risk) further diagnostic assessment is important when screening indicates that a child is likely to be at-risk. Evaluation in the following areas has been identified as important: intellectual, adaptive behaviour, personality/attitude, pre-academic, processing, language, speech, behaviour status, social/developmental background (Taylor, Willits, and Lieberman, 1990). Other authors emphasize the important distinction between screening and diagnosis (Meisels, 1987; Paget and Nagle, 1986; Satz and Fletcher, 1988). For policy development it is evident that structure and facility for using screening results in a responsible manner, through referring for comprehensive evaluation when screening results warrant, is of paramount importance.

Evaluation of Screening Policy

Consistent with the need to ensure validity of the measurements used, is the idea that any screening procedure used should be part of ongoing evaluation of assessment policy. Satz and Fletcher (1988) most concisely clarify what needs to be done with respect to validity studies for screening instruments. They point out the dearth of adequately conceptualized, designed, and implemented research

studies, noting that the poor research often conducted is not acceptable. They indicate that evaluation studies should incorporate:

- a) longitudinal design, b) independent assessments of kindergarten performance and learning ability separated by a temporal interval of at least three years, c) random sampling of children in a validation/cross validation design, and d) systematic assessment of predictive validity and utility. (p.829).

Recommendations

Given the overall trend of the results of this study which suggest that the MAP is of negligible value with respect to prediction of academic success, in and of itself, it is recommended that:

1. The school district reconsider the manner in which it uses the MAP as a kindergarten screening instrument. The evidence from this study provides no support for the use of the MAP as a predictor of grade one achievement. Therefore these results cannot be used to support the continued use of the MAP as a screening instrument. The school district would benefit from a planned policy with respect to a) what information they hope to obtain from screening, and b) the uses to which screening information is put.
2. The following considerations be included in planning and evaluation of screening policy.

(a) The school district should determine on the basis of its mission and mandate, the effect of screening on special education policy and practice. If screening is to be used for early identification purposes, then the types of special needs, and the percentage of special needs children whom the district intends to

service through special education, should be determined prior to the choice of screening instruments. A screening instrument (or any other assessment measure) can only be valid for a specific purpose. Until this purpose is identified and operationally defined the extent to which the MAP or any other instrument is valid for that purpose, cannot be adequately determined.

(b) Alternatively, the school district identify criterion related procedures for early identification of special needs children. With such criterion related considerations in place the extent to which the MAP is able to function as an adequate measure of the criterion can be evaluated. To the extent that the MAP is being used to direct programming or determine readiness for the E.C.S. curriculum, the school district should be aware that the instrument was designed as a developmental screening measure, and not a readiness test. Its psychometric properties, including concurrent, construct, and predictive validity, as well as its reliability, have not been proved sufficient for any such use of MAP results.

(c) The reliability of the MAP, its hit-rate, sensitivity, and specificity,, be further evaluated and such information be used in determining its adequacy for the purposes planned.

(d) Specific policy with respect to follow up comprehensive assessment on the basis of MAP results be implemented. Screening measures, by their nature will not identify special needs populations with adequate reliability for placement purposes. Cut-off scores which can be used to refer the appropriate percentage of this particular population for comprehensive diagnostic assessment will lead to defensible use of the MAP results.

Implications for Further Research

Further research needs to be undertaken in order to further evaluate the use of the Miller Assessment for Preschoolers for various purposes. With respect to kindergarten screening, further research into the interaction of child characteristics and the environment (perhaps both home and teaching) appears to be necessary before improvements in predictions of academic achievement are likely (Fedoruk, 1990). It is possible that assessment of student characteristics combined with observations and assessment of the ecological context will prove beneficial for aiding in identifying academic problems as early as possible.

Further statistical analysis which would examine characteristics such as the hit rate (false negatives and positives, true negatives and positives) of the MAP in identifying pre-academic problems could prove useful in the context of an ecological model of student success. Multivariate statistical techniques may be valuable in further improving the ability to predict achievement. Measures of student characteristics and behaviour, measures of teacher characteristics, and environmental characteristics, could be used in multiple regression techniques to develop the most powerful prediction equations.

Summary

The results of the study suggest that scores on the MAP are at best weakly to moderately correlated with scores on the Gates Reading Readiness Test, the Canadian Test of Basic Skills, and the Gates Reading Test, administered at various points throughout the grade one year. This chapter has presented a discussion concerning the possible factors relating to the weak correlations obtained. These include measurement and theoretical artifacts. The limitations of this study,

which relate to the design, sample, and data analysis employed were reviewed. Conclusions and recommendations included the suggestion that the school board reevaluate its use of the MAP, especially considering further research to determine how MAP results can best be implemented within a district policy incorporating, screening referral and programming.

References

- Academic Therapy Publications (1987)... *Quick Neurological Screening Test - Revised Edition*. Novato, CA: Author.
- American Psychological Association (1985). *Standards for Educational and Psychological Testing*. Washington: Author.
- Bracken, B. A. (1987). Limitations of preschool instruments and standards for minimal levels of technical adequacy. *Journal of Psychoeducational Assessment*, 4, 313-326.
- Bredenkamp, S. (Ed.) (1987) *Developmentally appropriate practice in early childhood programs serving children from birth through age 8*. Washington: National Association for the Education of Young Children
- Daniels, L. E. and Bressler, S. (1990). The Miller Assessment for Preschoolers: Clinical use with children with developmental delays. *American Journal of Occupational Therapy*, 44(1), 48-53.
- Drillien, C. M., Pickering, R.M., and Drummond, M.B. (1988). Predictive value of screening for different areas of development. *Developmental Medicine and Child Neurology*, 30, 294-305.
- Fedoruk, G. (1991). Kindergarten screening predictive inaccuracy: First grade teacher variability. *Exceptional Children*, 57(3), 258-263.
- Lehr, C. A., Ysseldyke, J.E., and Thurlow, M.L (1987). Assessment practices in model early childhood special education programs. *Psychology in the Schools*, 24, 390-399.
- McCarthy, D. (1972). *The McCarthy Scales of Children's Abilities*. San Antonio: TX: The Psychological Corporation.

- Meisels, S.J. (1987). Uses and abuses of developmental screening and school readiness testing. *Young Children*, January, 4-73.
- Miller, L.J. (1982). *Miller Assessment for Preschoolers*. Littleton, CO: The Foundation for Knowledge in Development: .
- Miller, L. J. (1984). *Overview of the Miller Assessment for Preschoolers*. International ALCD Conference. Pittsburgh, PA: ALCD.
- Miller, L. J. (1988). *Miller Assessment for Preschoolers Manual 1988 Revision*. Toronto: The Psychological Corporation, Harcourt Brace Jovanovich Inc.
- Nelson Canada Limited (1980). *Gaess-MacGinitie Reading Tests*. Nelson Canada Limited.
- Nelson Canada Limited (1981). *Canadian Test of Basic Skills*. Nelson Canada Limited.
- Nunnally, J. C. (1970). *Psychometric Theory*. New York: McGraw Hill.
- Padget, S.Y. (1985). Test review: Miller Assessment for Preschoolers. *Reading Teacher*, 39, 184-188.
- Paget, K.D. and Nagle, R.J. (1986). A conceptual model of preschool assessment. *School Psychology Review*, 15(2), 154-165.
- Provost, B., Harris, M.B., Ross, K., and Michnal, D. (1989). A comparison of scores on two preschool assessment tools: Implications for theory and practice. *Physical and Occupational Therapy in Pediatrics*, 8(4), 35-51.
- Salvia, J. and Ysseldyke, J.E. (1985). *Assessment in Special and Remedial Education* (3rd ed.). Boston: Houghton Mifflin.
- Sattler, J. M. (1988). *Assessment of Children*. San Diego, CA: Author.

- Satz, P. and Fletcher, J. M. (1988). Early identification of learning disabled children: An old problem revisited. *Journal of Consulting and Clinical Psychology, 56*(6), 824-829.
- Slaton, D.S. (1985). The Miller Assessment for Preschoolers: A clinicians perspective. *Physical and Occupational Therapy in Pediatrics, 5*(1), 65-70.
- Stevenson, J. (1984). Predictive value of speech and language screening. *Developmental Medicine and Child Neurology, 26*, 528-538.
- Taylor, R.L., Willits, P., and Lieberman, N. (1990). Identification of preschool children with mild handicaps: The importance of cooperative effort. *Childhood Education, Fall*, 26-31.
- Teachers College Press (1966). *Gates MacGinitie Reading Tests: Readiness Skills*. New York, NY: Teachers College Press.
- Wechsler, D. (1967). *Wechsler Preschool and Primary Scale of Intelligence*. San Antonio, TX: The Psychological Corporation.
- Wilson, B.J., and Reichmuth, M. (1985). Early-screening programs: When is predictive accuracy sufficient? *Learning Disability Quarterly, 8*, 182-188.