

EXAMINATION OF THE POTENTIAL OF SELECTED NORM-REFERENCED  
TESTS AND SELECTED LOCALLY DEVELOPED CRITERION-REFERENCED  
TESTS TO CLASSIFY STUDENTS INTO PERFORMANCE CATEGORIES

by

Renee Y. Jacobson

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Education

Major: Educational Administration

Under the Supervision of Professor Larry L. Dlugosh

Lincoln, Nebraska

May, 2008

UMI Number: 3303721

PREVIEW

UMI<sup>®</sup>

---

UMI Microform 3303721

Copyright 2008 by ProQuest Information and Learning Company.  
All rights reserved. This microform edition is protected against  
unauthorized copying under Title 17, United States Code.

---

ProQuest Information and Learning Company  
300 North Zeeb Road  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

Signature Page

PREVIEW

EXAMINATION OF THE POTENTIAL OF SELECTED NORM-REFERENCED  
TESTS AND SELECTED LOCALLY DEVELOPED CRITERION-REFERENCED  
TESTS TO CLASSIFY STUDENTS INTO PERFORMANCE CATEGORIES

Renee Y. Jacobson, Ed.D.

University of Nebraska, 2008

Adviser: Larry L. Dlugosh

The purpose of this study was to evaluate selected norm-referenced assessments (NRA) developed by achievement test publishers and selected criterion-referenced assessments (CRA) developed by teachers in Nebraska schools as of the spring of 2003. The study measured the sufficiency of the sampled assessments that aligned with selected Nebraska L.E.A.R.N.S. reading and math content standards to categorize student performance into two-level proficient or not proficient categories, and into four-level beginning, progressing, proficient, and advanced performance categories.

This study was an effort to evaluate the status of the CRA work-in-progress in the spring of 2003, and to make general sufficiency comparisons between the published NRAs and locally developed CRAs with regard to their capacity to assess student performance on selected Nebraska reading and mathematics content standards.

Norm-referenced and criterion-reference assessments evaluated in this study were able to categorize student performance levels for the selected reading and mathematics content standards. Published NRAs and locally-developed CRAs both had the capacity to distinguish student performance into proficient or not proficient. Both NRAs and CRAs were less sufficient at assessing four-level student performance.

NRAs and CRAs both sufficiently assessed student performance on reading content standards for the two-level performance categories (proficient or not proficient), but NRAs and CRAs were both less sufficient in assessing four-level student reading performance. In 2003 NRAs demonstrated stronger sufficiency in categorizing two-level student reading performance and somewhat stronger sufficiency in categorizing four-level student reading performance than CRAs demonstrated.

In 2003 student performance on mathematics content standards was assessed more sufficiently by NRAs for two-level performance at fourth grade and eighth grade. NRAs and CRAs were equivalent in assessing two-level student math performance for twelfth grade. Four-level math performance categorizations were more sufficiently assessed by NRAs for fourth grade mathematics. However, CRAS more sufficiently assessed four-level math performance at the eighth grade and twelfth grade.

PREVIEW

## ACKNOWLEDGEMENT

Major accomplishments require a supporting cast. This dissertation is the result of a team effort that included many professors, teachers and friends who encouraged me at each step of this journey. I will be forever grateful for the support.

This dissertation is dedicated:

...To my father who taught me that intelligence and education must be intertwined with commonsense.

...To my grandparents whose life lessons included an understanding that this life, though never perfect, provides great joy and abundant blessings.

...To my husband, children and mother whose love and encouragement continue to strengthen my mission to positively impact the lives of children today, tomorrow, and always.

How very blessed I have been! Thank you.

## CONTENTS

	Page
ACKNOWLEDGEMENT .....	i
LIST OF TABLES .....	iv
 CHAPTER	
I. INTRODUCTION .....	1
Purpose.....	5
Definitions.....	6
Research Questions .....	8
Methodology .....	9
II. LITERATURE REVIEW .....	12
History of Mental Measurement .....	12
Assessment in Elementary and Secondary Schools.....	13
Nebraska Assessment System.....	17
III. METHODOLOGY .....	25
Research Question .....	25
Methodology .....	25
Sample Selection.....	28
Sample of Assessments.....	28
Sample of Expert Evaluators .....	33
Selection of Content Standards .....	36
Procedure .....	37

IV. RESULTS .....	44
Phase One Results – Performance Standards and Definitions.....	45
Phase Two Ratings – Ratings of Assessment and Sufficiency .....	62
NRA Reading Evaluation Data.....	63
NRA Math Evaluation Data.....	65
CRA Reading Evaluation Data .....	68
CRA Math Evaluation Data.....	74
Summary of NRA and CRA Evaluation.....	81
Regional Consistency.....	84
Evaluation of Process by Participants.....	95
V. CONCLUSIONS AND RECOMMENDATIONS .....	103
Conclusions.....	105
Limitations and Recommendations.....	106
Recommendations for Next Steps.....	109
REFERENCES .....	111
APPENDIX .....	114
A – IRB .....	115
B – List of Schools with Exemplary or Very Good Ratings.....	119
C –Language Arts Standards Used in the Sufficiency Study .....	128
D –Mathematics Standards Used in the Sufficiency Study .....	129
E – Sample Letters and Permission Forms .....	130
F – Sample of Assessment Rating Sheet.....	140



G – Sample of Process Evaluation Form .....	145
H – Process Comments of Evaluators .....	148

PREVIEW

## LIST OF TABLES

TABLE	Page
1 Schools Providing CRA for Sufficiency Evaluation .....	32
2 Phase One Reading Evaluators Who Wrote Performance Definitions .....	34
3 Phase One Math Evaluators Who Wrote Performance Definitions.....	35
4 Phase Two Reading Evaluators Who Determined Difficulty of Assessment Tasks .....	36
5 Phase Two Mathematics Evaluators Who Determined Difficulty of Assessment Tasks .....	36
6 Board Performance Standards for Reading at 4 <sup>th</sup> , 8 <sup>th</sup> , and 12 <sup>th</sup> Grade Levels .....	45
7 Standard-Specific Operational Performance Definitions for Reading at 4 <sup>th</sup> , 8 <sup>th</sup> , and 12 <sup>th</sup> Grade Levels .....	47
8 Board Performance Standards for Mathematics at 4 <sup>th</sup> , 8 <sup>th</sup> , and 12 <sup>th</sup> Grade Levels .....	51
9 Standard-Specific Operational Performance Definitions for Selected Mathematics at 4 <sup>th</sup> , 8 <sup>th</sup> , and 12 <sup>th</sup> Grade Levels.....	52
10 NRA Consensus Ratings of Evaluators for Grade Four Nebraska L.E.A.R.N.S. Language Arts 4.1.3 .....	63
11 NRA Consensus Ratings of Evaluators for Grade Eight Nebraska L.E.A.R.N.S. Language Arts 8.1.1 .....	64
12 NRA Consensus Ratings of Evaluators for Grade Twelve Nebraska L.E.A.R.N.S. Language Arts 12.1.1 .....	64
13 NRA Consensus Ratings of Evaluators for Grade Four Nebraska L.E.A.R.N.S. Math 4.2.1 and 4.5.1.....	65
14 NRA Consensus Ratings of Evaluators for Grade Eight Nebraska L.E.A.R.N.S. Math 8.2.2 and 8.5.2.....	66
15 NRA Consensus Ratings of Evaluators for Grade Twelve Nebraska L.E.A.R.N.S. Math 12.2.1 and 12.5.1.....	67
16 CRA Consensus Ratings of Evaluators for Grade Four Nebraska L.E.A.R.N.S. Reading Standards .....	69

17	CRA Consensus Ratings of Evaluators for Grade Eight Nebraska L.E.A.R.N.S. Reading Standards .....	71
18	CRA Consensus Ratings of Evaluators for Twelfth Grade Nebraska L.E.A.R.N.S. Reading Standards.....	73
19	CRA Consensus Ratings of Evaluators for Grade Four Nebraska L.E.A.R.N.S. Selected Math Standards .....	76
20	CRA Consensus Ratings of Evaluators for Grade Eight Nebraska L.E.A.R.N.S. Selected Math Standards .....	78
21	CRA Consensus Ratings of Evaluators for Grade Twelve Nebraska L.E.A.R.N.S. Selected Math Standards .....	80
22	Percent Sufficiency for Selected Reading and Mathematics Content Standards in Selected Norm-Referenced Assessments and Criterion- Referenced Assessments.....	82
23	Frequency of Matching Task Performance Level Assignments Among Regions for Grade Four Nebraska L.E.A.R.N.S. Selected Reading Standards.....	84
24	Frequency of Matching Task Performance Level Assignments Among Regions for Grade Eight Nebraska L.E.A.R.N.S. Selected Reading Standards.....	85
25	Frequency of Matching Task Performance Level Assignments Among Regions for Grade Twelve Nebraska L.E.A.R.N.S. Selected Reading Standards.....	86
26	Frequency of Matching Task Performance Level Assignments Among Regions for Grade Four Nebraska L.E.A.R.N.S. Selected Mathematics Standards .....	87
27	Frequency of Matching Task Performance Level Assignments Among Regions for Grade Eight Nebraska L.E.A.R.N.S. Selected Mathematics Standards .....	88
28	Frequency of Matching Task Performance Level Assignments Among Regions for Grade Twelve Nebraska L.E.A.R.N.S. Selected Mathematics Standards .....	89
29	Magnitude and Direction of Task Performance Level Assignments Among Regions for Grade Four Nebraska L.E.A.R.N.S. Selected Reading Standards.....	92

30	Magnitude and Direction of Task Performance Level Assignments Among Regions for Grade Eight Nebraska L.E.A.R.N.S. Selected Reading Standards.....	92
31	Magnitude and Direction of Task Performance Level Assignments Among Regions for Grade Twelve Nebraska L.E.A.R.N.S. Selected Reading Standards.....	93
32	Magnitude and Direction of Task Performance Level Assignments Among Regions for Grade Four Nebraska L.E.A.R.N.S. Selected Mathematics Standards .....	93
33	Magnitude and Direction of Task Performance Level Assignments Among Regions for Grade Eight Nebraska L.E.A.R.N.S. Selected Mathematics Standards .....	94
34	Magnitude and Direction of Task Performance Level Assignments Among Regions for Grade Twelve Nebraska L.E.A.R.N.S. Selected Mathematics Standards .....	94
35	Participant Evaluation of the Orientation Process (mean rating) .....	96
36	Participant Evaluation of Orientation Time Allocation (mean rating) .....	97
37	Participant Evaluation of NRA Ratings and Time Allocation for NRA Ratings (mean rating) .....	98
38	Participant Evaluation of CRA Ratings and Time Allocation for CRA Ratings (mean rating).....	99
39	Participant Evaluation of Overall Success of the Study and Study Organization (mean rating).....	100

## **Chapter One**

### **Introduction**

Societal expectations of public education in the United States have evolved from education for the elite in the 1700s, to compulsory education in separate but equal schools in the 1800s and early 1900s, then to expectations of equal educational opportunities for all children in the last half of the 1900s. In 2003 societal expectations of public education, as delineated in federal legislation, commonly referred to as the No Child Left Behind Act, is proficient educational performance for every child. The paradigm has shifted the focus of school accountability away from providing educational opportunities for children to accountability for ensuring the learning outcomes for each child.

Measuring the degree to which public education meets societal expectations has evolved as well. In early 1996 the Nebraska Department of Education (NDE) began the process of constructing an assessment and accountability system for use in K-12 public schools. NDE personnel worked with groups of teachers and educators from across the state to develop core-area content standards. Their effort resulted in academic content standards in the core curricular areas of reading, writing, speaking, and listening; mathematics; science; and history and social studies for selected grade levels of 1<sup>st</sup>, 4<sup>th</sup>, 8<sup>th</sup> and 12<sup>th</sup>. These content standards were titled Leading Educational Achievement through Rigorous Nebraska Standards (Nebraska L.E.A.R.N.S.). Local school districts were encouraged to either adopt Nebraska LEARNs content standards or develop their own local standards. If a district chose to adopt local standards, the standards were required to

be of sufficient rigor and quality that the Nebraska Board of Education could approve them as equal to, or more rigorous than, the state's standards.

Despite a Unicameral mandate for a single state test in 1998, Commissioner of Education, Dr. Doug Christensen, appealed to Nebraska legislators' deeply rooted philosophy of local control and the growing body of research evidence that indicated significant negative consequences for students when high stakes state testing was implemented. Commissioner Christensen's testimony, the support of state educators, and other evidence convinced state legislators that the achievement of Nebraska students would be more likely to improve if classroom teachers were provided appropriate tools and an expectation to assess and report the status of student learning in their schools. Commissioner Christensen said, "Decisions about whether or not students are learning should not take place in the legislature, the governor's office, or the department of education. They should take place in the classroom because that is where learning occurs (Christensen, 2000)."

In addition to locally developed and administered assessments, Nebraska schools provided two other sources of assessment data. Schools were required to administer a standardized, norm-referenced achievement test of their own selection to students at least once in elementary, middle, and high school grades. The third source of assessment data from students of Nebraska schools was a statewide writing assessment at grades 4, 8, and 11. Together these three types of measurement, criterion-referenced assessment, norm-referenced assessment, and a state writing criterion-referenced assessment provided comprehensive, multidimensional information about student learning in Nebraska.

STARS is an acronym used to refer to this three-pronged assessment system. STARS stands for School-based, Teacher-led Assessment and Reporting System.

The original plan for STARS was a rotating system of reporting, i.e., reading and writing were to be reported in year one and mathematics in year two, continuing to alternate from then on. Science was to be added in year three, and Social Studies in year four. The state writing assessment was intended to annually alternate among 4<sup>th</sup>, 8<sup>th</sup> and 11<sup>th</sup> grade students. The first reporting requirement using this system was intended to be, and in fact was, reading and writing in 2000-01, and mathematics in 2001-02.

As with every new process, schools identified weaknesses in their systems that needed attention. At the urging of local school districts, NDE modified its planned implementation timeline to allow reading and writing assessments to be revised without adding science assessments in 2002-03, and likewise for mathematics without adding social studies in 2003-04. Science was rescheduled to come on line in 2004-05 and social studies in 2005-06. Thus, allowing schools to make needed changes in their assessments of language arts and mathematics before proceeding to science and social studies.

Meanwhile, on January 8, 2002 federal legislation was enacted mandating complex student proficiency testing for public schools. This legislation was commonly known as the No Child Left Behind Act of 2001 (NCLB). NCLB required states to report to the federal government the extent to which students in grades three through eight perform proficiently in reading comprehension and math problem solving using state-defined standards. Though the reporting requirement was a dichotomous classification, either proficient or not proficient, to be in full compliance for purposes of funding the legislation required students to be classified into four performance levels. Nebraska

named the performance levels “beginning”, “progressing”, “proficient”, and “advanced”.

Beginning and progressing levels were “not proficient” classifications. Proficient and advanced levels were “proficient” classifications. The Nebraska Department of Education personnel worked closely with the United States Department of Education personnel to preserve the integrity of the emerging state system of accountability (STARS), while meeting the federal requirements for reporting reading comprehension and math problem solving performance data each year for students in grades three through eight.

Regardless of the frequency (every year or in grades 4, 8 and 11) or the number of content areas (reading comprehension and math problem-solving, or reading, writing, mathematics, science, and social studies), it was essential that assessment tools were valid and reliable measures of student performance. Several conditions were needed to make valid, reliable classifications of student performance. First, each performance category must be clearly defined. In 2003 Nebraska school districts were responsible for developing both the conceptual performance standards (descriptions of what students should know and be able to do) and the operational performance definitions (assessment cut scores) for each performance category. Though NDE did not require specific performance standards, they did offer to provide exemplars of appropriate performance standards as a helpful guide to Nebraska school districts. Local school districts were under no obligation to adopt these performance standards because each school district was allowed to determine its own rigorous assessment system. Each district’s system was to be constructed in accordance with educational measurement and assessment industry standards for quality and rigor.



Second, NDE required each school district to insure that their assessments contained tasks that were sufficient in quantity and quality to permit students to be validly and reliably classified into each of the four performance levels. For example, if all tasks used to assess a particular standard were appropriate only for measuring the skill levels of beginning students, then students who possess advanced skill levels would not have an opportunity to demonstrate their skill level. Similarly, if all assessment tasks required advanced-level performance, students who had not yet reached the advanced performance level would be unable to demonstrate their developing skill level. The only inference that could be made about the students who had not reached the advanced level was that they were not advanced. This information would not be helpful in making distinctions among beginning, progressing and proficient students. In addition to having questions at each performance level, the number of questions at each level must be of a sufficient number to reliably classify the student's performance on a given content standard. Too many questions would waste student learning time, but too few questions render the student performance data unreliable.

#### *Purpose*

The purpose of this study was to evaluate the extent to which in the summer of 2003 Nebraska K-12 public school districts had assessments that sufficiently assessed the four levels of performance. This study was needed to determine next steps for the Nebraska Department of Education and Nebraska schools. The study examined locally developed criterion-referenced assessments and commonly used norm-referenced assessments.

#### *Definitions*

Listed below are many terms used in this dissertation and their meaning as they have been used in the context of this document.

*Content Standard.* An expectation of what a student at a particular grade level should know and be able to do within an academic content area such as math, language arts, science or social studies, defines expected student skills and knowledge.

*Leading Educational Achievement through Rigorous Nebraska Standards (Nebraska L.E.A.R.N.S.)* Content standards developed by Nebraska educators and adopted by the Nebraska Board of Education and Nebraska Legislature. The standards address language arts (reading, writing, speaking, and listening), mathematics, science, and social studies. School districts were allowed to adopt L.E.A.R.N.S. standards or develop standards that were equal to or more rigorous than the state standards (Nebraska Department of Education, 2004).

*Nebraska Department of Education (NDE).* The Nebraska Constitution in article seven, section two provides for the establishment of the state department of education. The Reissued Revised Statutes of Nebraska §32-511 defines the NDE as consisting of a state board of education and a commissioner of education. The eight board members are elected by district on a nonpartisan ballot at the statewide general election. Each board member serves a four-year term. The commissioner of education is appointed by the state board of education. Reissue Revised Statutes of Nebraska §79-301, section 1 clarifies the role of NDE as providing for the general supervision and administration of the school system of the state.

*No Child Left Behind Act of 2001 (NCLB).* A federal act signed into law on January 8, 2002 by President George W. Bush requiring states to implement an accountability

system that includes annual state assessments of student performance on state academic achievement standards. The act also sets requirements for teacher quality and public reporting (Public Law 107-110, 107<sup>th</sup> Congress, 20 USC 6301 Improving the Academic Achievement of the Disadvantaged, 2001).

*Performance Standard.* A performance level that demonstrates proficiency at specific levels in content standards (Wright, 1998).

*Response Probability (RP).* In assessment item classification response probability is the percent of students who are expected to respond accurately to a test item for it to be considered appropriate for that level, e.g., if a question is defined as progressing, one would expect 67 percent of progressing level students to answer it correctly, a smaller percentage of beginning level students would respond correctly, and a greater percentage of proficient and advanced students would respond correctly. A RP of 67 percent is commonly used by the psychometric community as recommended by Lewis, et. Al. (1996).

*School-based Teacher-led Assessment and Reporting System (STARS).* Nebraska's three-pronged system of educational accountability for K-12 school districts. The system includes criterion-referenced assessments that were teacher-designed and teacher-administered assessments that meet six quality criteria. This assessment component focuses on direct measurement of student performance on content standards. In addition, schools participate in a statewide writing assessment, and schools administer norm-referenced assessments of academic achievement (NDE, 2004).

*Sufficiency.* Individuals who are assessed are provided appropriate opportunities in adequate quantity at each performance level to provided valid and reliable student performance judgments.

*Unicameral.* The one-house, nonpartisan legislative body of state government adopted in 1937 by Nebraska voters to cut government costs by replacing the two-house legislative body of state government (Nebraska Unicameral, 2006).

The reader will observe the use of the terms “assessment” and “test” throughout this document. In this document an assessment refers either to the collection of tests used by an organization to determine the performance of students or all assessments that align with a given content standard. A test shall refer to all tasks that a student would see during a testing event. For example, a norm-referenced assessment might include tests of reading, spelling, math, social studies, and science; and the reading, social studies, and science tests might each contain items that align with a reading comprehension content standard. In the course of this research project the investigators looked at the test items and tasks of various assessment systems that aligned with selected content standards.

### *Research Question*

The research question used to focus this study was: Do selected norm-referenced tests and selected locally developed criterion-referenced tests as they existed in the spring of 2003 have the capacity to sufficiently categorize student performance on selected reading and math standards in Nebraska?

### *Methodology*

Assessment sufficiency requires that students taking a “test” have quality opportunities in adequate quantity at each performance level to provide valid and reliable student performance judgments. Two basic types of methodologies prevail in educational psychometrics for evaluating the sufficiency of assessment instruments. One method relies upon the collection and comparison of correlated data. This method involves comparing performance data produced by an assessment instrument to a source of performance data that is generally accepted to be a valid and reliable proxy of student performance. The second method uses a panel of experts to compare each component of an assessment instrument to a constructed performance standard (Messick, 1989; Scriven, 1967). The method selected for this study relied upon the second methodology using panels of experts.

The expert panel method was selected because in 2003 there is no “generally accepted” data available to categorize student performance as beginning, progressing, proficient and advanced, particularly as student performance relates to Nebraska’s content standards. Federal legislation in NCLB allowed each state to define the four levels of performance for reading and math problem solving, just as Nebraska allowed each local district to define each of the four performance categories. As such, there was no generally accepted student performance data available to which data from NRAs or locally developed CRAs could be correlated.

For an expert panel to evaluate the assessment tasks associated with a content standard, they needed a performance level standard and operational definitions to which each assessment task could be compared. Because neither federal nor state level performance standards were in existence at the initiation of this study, the first step of the

study was to develop them. Once performance standards and operational definitions were developed during the summer of 2003, panels of experts were convened in the fall of 2003 to evaluate selected assessments. The experts were trained to work as a team to determine and record the performance difficulty of each task from selected test instruments. Using the data produced by the expert panels, the investigators analyzed the data to conclude the sufficiency of each assessment instrument.

The following general procedures guided the process which was conducted during the summer and fall of 2003.

1. A representative sample of assessment literate Nebraska educators reached consensus on and committed to writing broad conceptual definitions of performance levels (beginning, progressing, proficient, and advanced) for selected reading standards at each of three grade levels, 4<sup>th</sup>, 8<sup>th</sup> and 11<sup>th</sup>.
2. A representative sample of assessment literate Nebraska educators reached consensus on conceptual definitions of performance levels (beginning, progressing, proficient, and advanced) for selected mathematics standards at each of three grade levels, 4<sup>th</sup>, 8<sup>th</sup> and 11<sup>th</sup>.
3. Using the performance level definitions developed in steps one and two, a representative sample of assessment literate Nebraska educators determined the potential performance level of individual assessment tasks from commercially developed norm-referenced assessments.
4. Using the performance level definitions developed in steps one and two, a representative sample of assessment literate Nebraska educators determined the

potential performance level of individual assessment tasks on locally-developed, criterion-referenced assessments.

5. Investigators evaluated the range and quantity of NRA and CRA test items to determine the potential of selected NRAs and CRAs to categorize student performance as beginning, progressing, proficient or advanced for selected content standards.

This project was conducted as a joint project with the Buros Center for Testing, specifically the Buros Institute for Assessment Consultation and Outreach (BIACO) and the Nebraska Assessment and Accountability Comprehensive Evaluation Project. An Institutional Review Board (IRB) application was submitted and approved by the Nebraska Assessment and Accountability Comprehensive Evaluation Project. The author was named as an investigator on that application. A copy of the IRB approval is attached in Appendix A.

## **Chapter Two**

### **Literature Review**

#### *History of Mental Measurement*

Mental measurements have been in existence for many years. The ancient Greek philosophers Socrates, Plato and Aristotle contemplated the origins of intelligence and the influence of heredity and environment (Menon, 1956). From the 7<sup>th</sup> century to the early 20<sup>th</sup> century the Chinese used a civil service examination to determine the “moral behavior” of men. Moral behavior was obtainable through extensive education according to Confucian philosophy (Lee, 1985). Similarly, Britain and the United States have used examinations to identify candidates for specific jobs. From 1870 to 1925, British civil service examinations tested a candidate’s writing ability, and was a good predictor of the ability to manage large quantities of paperwork under stressful conditions (Russell-Smith, 1974). At about the same time the United States implemented a system of testing army recruits to help make more efficient job assignments (Kevles, 1968).

At the turn of the twentieth century high schools in the United States were encouraged by business and industry to add vocational coursework thereby expanding their curricula beyond only classical studies. Concurrently, colleges began to use educational testing that was specific to their institution to determine which students would be accepted into the finest schools and colleges. This testing was based mostly upon knowledge of the classic curriculum of Latin, Greek, philosophy, rhetoric and mathematics. The need for graduate colleges to develop a system of selection to ensure a students’ academic ability arose when high schools became more diverse in their offerings. The cost and logistics involved in managing and maintaining a relevant