Iowa Algebra Readiness Assessment

Manual for Test Use, Interpretation, and Technical Support

Iowa Testing Programs

Prepared at The University of Iowa by

Harold L. Schoen, Timothy N. Ansley

With the assistance of

Bradley Thiessen, Judy Feil, Michelle Mengeling, and H. D. Hoover

Table of Contents

Overview

The *Iowa Algebra Readiness Assessment*™ *(IARA*™*)* was developed to help teachers and counselors make the most informed decisions possible regarding the initial placement of students in the secondary mathematics curriculum. Because the stakes associated with this process tend to be high, it is important that those making the decisions can supplement their professional judgment with objective information about the mathematical abilities of students.

The *IARA* provides this kind of objective information. Designed to gather information efficiently, the test provides a four-part student profile that facilitates the diagnosis of strengths and weaknesses. In addition, the test provides information across multiple domains of the Common Core State Standards in Mathematics. Because time is always at a premium the test requires only a typical class period.

Test Purposes

Given the increasing importance of readiness in mathematics in today's technological world, it is essential that students not be placed in mathematics courses for which they are not prepared; nor should they end up in courses in which they will be unchallenged and bored. In either case, students tend to drop out of the mathematics curriculum before they have a fair chance to evaluate their interest and aptitude in this area. One of the main objectives of education is to open doors of opportunity to students. Too often, the door to mathematically oriented careers is closed far too early.

In making decisions regarding mathematics course placement, the recommendations of teachers must be given greatest weight. However, they cannot be the only criterion. Junior high or middle school students will typically have had many different teachers, and it is unlikely that they all share common standards for judging students' mathematical abilities. Students moving into a district raise another problem. To address these issues, some schools have elected to use scores from standardized achievement tests to aid in making placement decisions for students. Usually these tests have not been validated for this purpose, and it is therefore not wise to use them in this manner. Other schools have created their own local screening tests. Yet if such instruments are to be used with confidence, they should meet the same criteria that a standardized test does in terms of evidence of validity and precision of measurement.

There is a need for a standardized measure specifically constructed to indicate mathematics readiness to assist in placement decisions, and it is for this purpose that the *IARA* was developed.

Background and Rationale for the *IARA*

As background for guiding the *IARA,* a number of sources were examined: the treatment of pre-algebraic and algebraic topics in middle school curricula; position papers by mathematics educators suggesting what students should learn in pre-algebra and algebra; the current state of the curriculum and teaching of algebra in high schools; research articles on students' learning of basic algebra concepts and their development of symbol sense; and the *Principles and Standards for School Mathematics* from the National Council of Teachers of Mathematics. Various themes emerged from this investigation, including the generalization of patterns, pre-algebra skills, algorithmic thinking, and linking multiple representations of algebraic concepts and relationships. The *IARA* consists of the following four content categories:

Pre-Algebraic Number Skills and Concepts (15 items)

Measures how well students understand some of the mathematical skills and concepts necessary to be prepared for a course in algebra. The first several questions assess students' skills with integer arithmetic. The remaining items measure conceptual understanding and problem-solving skills.

Interpreting Mathematical Information (15 items)

Includes graphs and verbal/symbolic definitions of mathematical concepts followed by several questions about the material. This part assesses how well a student can learn new material presented graphically or textually.

Representing Relationships (15 items)

The first five items present relationships between two sets of numbers in a table, and the students must find the general rule for the relationship. These rules involve all four arithmetic operations and positive and negative numbers. The remaining items measure how well students interpret and represent relationships from information presented in verbal, graphic, or symbolic form.

Using Symbols (15 items)

Items address common misconceptions about variables, equations, removing parentheses, consecutive integers, and variation of one number in an expression when others are held constant.

Common Core State Standards

The materials have also been aligned to the Common Core State Standards (CCSS) in Mathematics. The six domains of the CCSS that are sampled are:

- Expressions and Equations
- Statistics and Probability
- Ratio and Proportional Relationships
- The Number System
- Geometry
- Functions

Table 1 contains the test specifications for the *IARA.*

Table 1
***Iowa Algebra Readiness Assessment* Table of Specifications**

Content Categories	Items
Pre-Algebraic Number Skills and Concepts	**1–15**
Using Numerical Expressions	1, 2, 3, 4, 5
Solving Problems	
Simple	9, 10, 11, 12
Complex	6, 7, 8, 13, 14, 15
Interpreting Mathematical Information	**16–30**
Graphs	
Single Points	16, 18, 25
Graph as a Whole	17, 19, 24, 26, 27
Novel Mathematical Formulations	
Comprehension	20, 21, 28
Application of Symbols	22, 23, 29, 30
Representing Relationships	**31–45**
Inferring Functional Relationships	31, 32, 33, 34, 35, 36, 37
Expressions with Variables	38, 39, 40, 41, 42, 43, 44, 45
Using Symbols	**46–60**
Solving Algebraic Expressions	46, 48, 49, 50, 51, 53, 54
Applying Symbolic Representation	47, 52, 55
Identifying Relationships Among Variables in a Formula	56, 57, 58, 59, 60

Interpretation and Use of *IARA* Scores

Description of *IARA* Scores

The *IARA* provides scores on each of the four content categories, the composite score, and each of the five domains of the Common Core State Standards in Mathematics. The *IARA* provides four types of scores in addition to raw scores: standard scores, percentile ranks, stanines, and normal curve equivalents. The characteristics of these four scores are provided below.

Standard Scores

A raw score, or number correct score, on any test has limited inherent meaning. For this reason, the raw scores of most standardized tests are converted to a standard score (SS) scale. Standard score scales provide scores that are more directly interpretable than raw scores. Many different types of standard scores are used in practice. For example, many “intelligence” tests use a standard score scale with a mean of 100 and a standard deviation of 15. Such scales provide very convenient interpretations. The composite raw scores of the *IARA* are transformed so that the mean of these scores corresponds to a normalized standard score of 150, and the standard deviation of the resulting standard score distribution is 15. This composite raw score is merely the sum of the four content category raw scores. An examination of the distributions of the content category raw scores indicated that these four distributions were very comparable, especially in terms of variability. It was therefore reasonable to form a composite raw score simply as the sum of the unweighted content category scores. From the reliability and validity evidence gathered to date, decisions regarding mathematics course placement should be based on the composite scores. Therefore, only the composite scores are placed on the standard score scale.

Percentile Ranks

A second type of derived score used with the *IARA* scores is the percentile rank (PR). Percentile ranks are probably the most easily understood derived scores. For example, if Sue has a percentile rank of 77 on the *IARA* composite score, that simply means that Sue’s composite score was higher than 77% of those in the reference group with whom Sue is being compared.

Percentile ranks are provided for the four content categories as well as the composite. Given the ease with which PRs can be interpreted, it is strongly recommended that PRs be used extensively in assisting parents and students to interpret individual scores as well as make comparisons among the four content categories. It should be noted, however, that some caution is necessary in considering or comparing PRs, especially when these values are near the middle of the distribution. Differences in PRs near the median are not as significant as differences of similar magnitude in the extremes of the distribution. For example, the difference between the 50th and 55th percentiles is less significant than the difference between the 90th and 95th percentiles. This is due to the high concentration of scores near the middle of the distribution and the relatively low concentration of scores at the extremes.

Stanines

Stanine scores can also be computed for each *IARA* content category and the composite scores. A stanine score is a normalized standard score with a range from 1 to 9 and an average of 5. Like percentile ranks, stanines are status scores with respect to the norm group. Generally, stanines may be considered to correspond to the following groupings of percentile ranks:

Percentile Ranks	Stanines
96–99	9
89–95	8
77–88	7
60–76	6
41–59	5
24–40	4
12–23	3
5–11	2
1–4	1

Stanine scores have the advantages of being readily interpretable and convenient, since they consist of only a single digit. Stanines also can serve to minimize the importance of very small score differences. However, at best stanines provide a very rough interpretation of student performance, as evidenced by the fact that two students achieving at the 59th and 60th percentiles will be assigned different stanines, while students at the 40th and 59th percentiles will be assigned the same stanine.

Normal Curve Equivalents

Normal curve equivalents (NCEs) are normalized standard scores with a mean of 50 and a standard deviation of 21.06. NCEs range from 1 to 99, corresponding to percentile ranks of 1.0 and 99.0, respectively. NCEs are not very informative to the typical test user. To interpret NCEs, it is typically necessary to compare them to some other status scores based on a common normative group, such as percentile ranks or stanines. For those users familiar with stanines, NCEs are often considered to be equivalent to stanines to one decimal place. For example, an NCE of 81 can be roughly approximated as a stanine of 8.1. The relationships among NCEs, percentile ranks, and stanines are illustrated in Table 2.

Table 2
Relationship of NCEs to PRs and Stanines

NCE	PR	NCE	Stanine
99	99	86–99	9
90	97	76–85	8
80	92	66–75	7
70	83	56–65	6
60	68	45–55	5
50	50	35–44	4
40	32	25–34	3
30	17	15–24	2
20	8	1–14	1
10	3		
1	1		

Interpretation of *IARA* Scores

In interpreting *IARA* scores, it is essential that teachers, counselors, parents, and students clearly understand that these scores are only meaningful in combination with other information available regarding the mathematical abilities of students. These test scores should serve as only one piece of the puzzle in determining where to place students in the secondary mathematics curriculum. While the *IARA* scores provide some unique, objective information regarding the mathematical abilities of students, these scores can in no way substitute for the professional judgments of teachers and counselors. In nearly all cases, the data obtained from the *IARA* should be less important than the observations and recommendations of teachers who have direct knowledge of the students. Also of obvious importance in making such decisions are the personal characteristics and goals of the students. These factors considered in conjunction with the *IARA* scores can provide an accurate estimate of students' readiness to study algebra.

The four *IARA* content category scores and the Common Core scores can provide some diagnostic information in the evaluation of a student's mathematical abilities. Students should be given a record of these scores to be shared with their parents. The four content category scores and he Common Core scores represent a profile of the mathematical skills of a student and should be so interpreted. Unusually strong or weak performance on any of the categories might merit further investigation. In most cases, the *IARA* scores will serve to reinforce teachers' perceptions. However, in some instances, the profile might yield valuable and unique insights into the depth and breadth of a student's facility with mathematical reasoning. If, for example, a student's percentile rank for "Interpreting Mathematical Information" was 33, while other content category percentile ranks were above 80, a teacher would probably wish to determine potential causes for this relative weakness. If a teacher believes that the content measured in "Interpreting Mathematical Information" is important, he or she might decide to provide remediation for this student in this area. As always, test scores are just a single piece of information and are not completely free from error. As such, any interpretation of test scores must be made in the larger context of all of the information available regarding a student's abilities.

Teachers should receive a listing of the performance of the students in their classes. These scores can be examined individually as described above, or they can be considered aggregately at a class or group level. For example, it might be of interest to identify content categories for which the class as a whole is relatively weak. This may have implications for instruction. If desired, group means for each content category could be computed. These average scores could be assigned percentile ranks from the conversion tables in the Appendix. These percentile ranks would reflect the performance level of a typical student in the group of interest.

Test Difficulty

No single test can be perfectly suited in difficulty for all students in a heterogeneous grade group. To maximize the reliability of a ranking within a group, a test such as the *IARA* must utilize nearly the entire range of possible scores; the raw scores on the test should range from zero to the highest possible score. In order to insure such a spread, test developers generally conduct one or more preliminary tryouts of items in order to objectively determine the relative difficulty and discriminating power of the items. A few of the items in the test should be relatively easy so that most students answer them correctly. Similarly, a few very difficult items are included to challenge the most able students. Most items, however, should be of medium difficulty and should discriminate well at all levels of ability. In other words, the typical student will succeed on only a little more than half of the test items, while the least able students may succeed on only a few of them. A test constructed in this manner results in the widest possible range of scores and yields the highest reliability per unit of testing time.

The *Iowa Algebra Readiness Assessment* was constructed to discriminate in this manner among students who may be eligible to enroll in a first course in Algebra. A summary of the distributions of item difficulties is given in Table 3. These distributions come from the 2005 national standardization sample. As can be seen from these statistics, the overall goal of obtaining an average item difficulty in the neighborhood of .40 – .60 was met. Tests with average item difficulty levels in the vicinity of .50 will have optimal internal consistency reliability (Ebel & Frisbie, 1991).

These data also illustrate the variability of item difficulty needed to discriminate throughout the entire ability range. It is extremely important in test development to have both relatively easy and relatively difficult items. Not only are such items needed for motivational reasons, but also they are critical if a test is to have enough ceiling for the most capable students and enough floor for the least capable students. There are items on the *IARA* with difficulties in the .70s and .80s, as well as some in the .20s and .30s.

It should be noted that these difficulty characteristics are for a cross section of the attendance centers in the nation. The distributions of item difficulties vary markedly among attendance centers, both within and between school systems.

Discrimination

Table 3 also presents a summary of the discrimination indices (item-test biserial correlations). Item discrimination indices are routinely used in item tryouts as one of the criteria for item selection. Table 3 provides the means, medians, and the 90th and 10th percentiles in the distributions of biserial correlations. As would be expected, discrimination indices vary considerably. Higher values of these indices are desirable. The values of these correlations can be limited in association with the item difficulties. For example, it is unusual to have very high discrimination values for very hard or easy items.

Ceiling Effects, Floor Effects, and Completion Rates

A summary of ceiling and floor effects is shown in Table 4. On the top line of the table is the number of items (k) in each test. Under "Ceiling," the percentile rank of a perfect score is listed as well as the percentile rank of a score one less than perfect (k – 1).

A "chance" score is frequently defined as the number of items in the test divided by the number of response options per item. The percentile ranks of these "chance" estimates are listed under "Floor." Of course, not all students who score at this level do so by chance. However, when a substantial proportion of students score at this level, it is an indication that the test may be too difficult.

The data in Table 4 illustrate that the numbers of students achieving perfect scores or scores below chance were small.

Table 3
Summary of Item Difficulty and Discrimination Indices for the *IARA*
(2005 National Standardization Sample)

Test	Content Category 1	Content Category 2	Content Category 3	Content Category 4
Number of Items	**15**	**15**	**15**	**15**
	Difficulty			
Mean	0.52	0.50	0.53	0.59
Median	0.48	0.46	0.53	0.60
P_{90}	0.71	0.73	0.70	0.72
P_{10}	0.36	0.36	0.36	0.47
	Discrimination			
Mean	0.61	0.56	0.61	0.65
Median	0.62	0.59	0.64	0.65
P_{90}	0.78	0.75	0.72	0.75
P_{10}	0.46	0.42	0.45	0.52

Table 4
Ceiling Effects, Floor Effects, and Completion Rates for the *IARA*
(2005 National Standardization Sample)

Test	Content Category 1	Content Category 2	Content Category 3	Content Category 4
Number of Items (k)	**15**	**15**	**15**	**15**
	Ceiling			
PR of k*	99	99	99	99
PR of k–1	98	98	98	92
	Floor			
%<k/n**	7	4	7	8

*k = perfect raw score
**Floor = percent of students scoring below k/n, where n=number of response options

This section details validity and reliability evidence for the *IARA*.

Validity of the *IARA*

The validity of any measurement instrument is not an absolute characteristic; that is, the degree to which a test is valid varies depending on the intended use. Thus, test validation must be an ongoing process. The evidence provided here should serve the most general purposes of the *IARA*. In many instances, it will be beneficial or even necessary for test users to carry out their own validation studies according to the manner in which they plan to use their test scores. For example, the *IARA* might be used in junior or community colleges with adults needing to be placed in mathematics courses. Or, some schools might well want to test students in sixth grade. However, the normative and validation data presently available come only from students in grades 7 and 8. Local validation studies would need to be conducted to determine the appropriate score interpretation for other populations.

The main types of validity evidence necessary for tests used for selection or placement are *content validity* and *criterion-related validity*.

Content Validity

Content validity involves the adequacy and appropriateness of the nature and scope of the test items. Content validity is generally assessed subjectively by a thorough examination of the test content and careful consideration of the nature and scope of the items and objectives. (See Table 1 for *IARA* content specifications.) Extreme care was taken in the construction of the *IARA*. A rigorous item development plan was followed. This plan included a close examination of algebra and pre-algebra textbooks. Also, the mathematics education research literature was thoroughly examined to determine the current thinking on the beginning of the secondary mathematics curriculum as well as possible promising future directions. In addition, the NCTM *Standards* were a guiding force in the development of the *IARA*. Once this foundation was prepared, items were written, tried out, and revised. Following this regimen enabled the development of the most efficient and valid assessment device possible. Once constructed, the *IARA* was scrutinized by mathematics educators to evaluate the content validity of these tests. The suggestions of these content experts were incorporated into the final forms of the *IARA*.

It is important to note that in spite of the painstaking care used in the development of the *IARA*, each user should scrutinize the content of this test to determine its appropriateness for specific applications.

Fairness and Bias

Any standardized measure must take steps to ensure that the test scores are fair to all examinees and free from bias (AERA, APA, & NCME, 1999). The *IARA* was evaluated by educators for fairness and cultural sensitivity issues as well as for balance in regional, urban-rural, and male-female representativeness. These educators were selected to represent Native Americans, Asian Americans, African Americans, Hispanic Americans, and whites. Some items were revised or replaced based on the reviewers' suggestions. In addition, a statistical analysis of Differential Item Functioning (DIF) was carried out. This procedure aims to control for response differences between focal and reference groups that may be due to school curriculum and environment. This analysis flagged one item for DIF. This item was subsequently reviewed; no bias was discerned.

Criterion-related Validity

Criterion-related validity refers to the relationship between test scores and other measures of student achievement. When these other measures are obtained at approximately the same time as the test scores, the evidence gathered is referred to as *concurrent validity*. When the measures are obtained at a later time and the data reflect the accuracy of the test as a predictive tool, the evidence is referred to as *predictive validity*.

Typically, in attempting to use the *IARA* scores to assist in student placement, it will be necessary to derive some sort of "cut-score." This score must be chosen at the local school or district level and will likely vary from district to district. To choose a cut-score wisely requires careful consideration of local conditions. One matter of interest would be how the local norm group compares with the national norm group. Each district should begin to aggregate its *IARA* data together with students' subsequent mathematics course grades. Over time, a district can accumulate enough information regarding the relationship between *IARA* scores and performance in Algebra I to make very well-informed placement decisions for students about to enter the secondary mathematics curriculum. It is extremely important that a test used for selection purposes have evidence of criterion-related validity. A detailed discussion of a study (Barron, Ansley, and Hoover, 1991) conducted to gather criterion-related validity evidence is given in the Appendix.

Reliability of the *IARA* Scores

The reliability of a test refers to the accuracy with which the test measures the construct of interest. As is true for validity evidence, reliability evidence can be presented in several ways. Perhaps the most common involves the assessment of internal consistency of the item responses. The Kuder-Richardson Formula 20 (KR-20) reliability of the *IARA*, derived from the responses of the standardization sample, is 0.93.

APPENDIX

How to Convert Raw Scores to Percentile Ranks, Normal Curve Equivalents, and Stanine Scores.

To convert a **composite (total)** raw score to a composite (total) percentile rank (PR) score, normal curve equivalent (NCE) score, or stanine score, first use **Table A1** to convert the total raw score to a standard score (SS), making sure to use the data for the correct form of the test. Then using the SS you obtained, use **Table A2** to convert the SS to PR, NCE, and stanine scores. The SS to PR, NCE, and stanine conversions are the same for both forms.

To convert the raw score (RS) for a **content category** to a percentile rank (PR), first identify the form of the test taken and select the correct table for the corresponding content category and test form from **Tables A3–A6.** Then locate the raw score on the left-hand side and read across to the right.

Table A1
Raw Score to Standard Score Conversion Table for Composite Scores

Raw Score	Standard Score	Raw Score	Standard Score
0	100	**31**	150
1	101	**32**	151
2	102	**33**	152
3	103	**34**	153
4	104	**35**	154
5	106	**36**	155
6	108	**37**	156
7	111	**38**	157
8	114	**39**	158
9	117	**40**	159
10	119	**41**	160
11	121	**42**	161
12	123	**43**	162
13	125	**44**	163
14	128	**45**	164
15	130	**46**	165
16	132	**47**	166
17	133	**48**	167
18	135	**49**	169
19	136	**50**	170
20	138	**51**	172
21	139	**52**	174
22	141	**53**	176
23	142	**54**	178
24	143	**55**	180
25	144	**56**	182
26	145	**57**	185
27	146	**58**	189
28	147	**59**	194
29	148	**60**	200
30	149		

Table A2
Standard Score (SS) to Percentile Rank (PR) to Normal Curve Equivalent (NCE) to Stanine Conversions

SS	PR	NCE	Stanine
100–117	1	1	1
118	2	7	1
119	2	7	1
120	2	7	1
121	3	10	1
122	3	10	1
123	4	13	1
124	4	13	1
125	5	15	2
126	5	15	2
127	6	17	2
128	7	19	2
129	8	20	2
130	9	22	2
131	10	23	2
132	12	25	3
133	13	26	3
134	14	27	3
135	16	29	3
136	18	31	3
137	19	32	3
138	21	33	3
139	23	34	3
140	25	36	4
141	27	37	4
142	30	39	4
143	32	40	4
144	34	41	4
145	37	43	4
146	39	44	4
147	42	46	5
148	45	47	5
149	47	48	5
150	50	50	5
151	53	52	5
152	55	53	5
153	58	54	5
154	61	56	6
155	63	57	6
156	66	59	6
157	68	60	6
158	70	61	6
159	73	63	6
160	75	64	6
161	77	66	7
162	79	67	7
163	81	68	7
164	82	69	7
165	84	71	7
166	86	73	7
167	87	74	7
168	88	75	7
169	90	77	8
170	91	78	8
171	92	80	8
172	93	81	8
173	94	83	8
174	95	85	8
175	95	85	8
176	96	87	9
177	96	87	9
178	97	90	9
179	97	90	9
180	98	93	9
181	98	93	9
182	98	93	9
183–200	99	99	9

Table A3
Raw Score (RS) to Percentile Rank (PR) Conversions for Content Category 1

RS	PR
0	1
1	3
2	7
3	12
4	19
5	29
6	39
7	48
8	57
9	65
10	73
11	81
12	88
13	94
14	98
15	99

Table A4
Raw Score (RS) to Percentile Rank (PR) Conversions for Content Category 2

RS	PR
0	1
1	2
2	4
3	10
4	19
5	30
6	42
7	53
8	63
9	72
10	80
11	86
12	91
13	95
14	98
15	99

Table A5
Raw Score (RS) to Percentile Rank (PR) Conversions for Content Category 3

RS	PR
0	2
1	4
2	7
3	13
4	21
5	30
6	38
7	46
8	56
9	65
10	73
11	80
12	87
13	93
14	98
15	99

Table A6
Raw Score (RS) to Percentile Rank (PR) Conversions for Content Category 4

RS	PR
0	4
1	6
2	8
3	12
4	18
5	24
6	31
7	38
8	44
9	51
10	58
11	65
12	74
13	83
14	92
15	99

Table A7
Item Characteristics
(Based on the 2005 National Standardization Sample, Grades 7 & 8 Combined)

	Item #	Difficulty	Discrimination
Content Category 1			
	1	.63	.58
	2	.79	.68
	3	.60	.51
	4	.42	.56
	5	.49	.41
	6	.61	.42
	7	.67	.56
	8	.71	.44
	9	.48	.56
	10	.48	.60
	11	.44	.65
	12	.59	.64
	13	.37	.34
	14	.38	.48
	15	.30	.34
Content Category 2			
	1	.69	.37
	2	.56	.35
	3	.73	.44
	4	.82	.50
	5	.60	.52
	6	.39	.52
	7	.46	.51
	8	.37	.24
	9	.42	.53
	10	.57	.69
	11	.37	.49
	12	.48	.29
	13	.43	.68
	14	.37	.55
	15	.33	.55

	Item #	Difficulty	Discrimination
Content Category 3			
	1	.78	.61
	2	.46	.45
	3	.61	.59
	4	.39	.62
	5	.37	.58
	6	.59	.69
	7	.63	.65
	8	.70	.56
	9	.42	.54
	10	.46	.41
	11	.65	.63
	12	.53	.61
	13	.53	.62
	14	.56	.62
	15	.27	.44
Content Category 4			
	1	.71	.64
	2	.54	.66
	3	.73	.70
	4	.69	.67
	5	.69	.67
	6	.51	.61
	7	.49	.59
	8	.60	.59
	9	.63	.77
	10	.40	.62
	11	.65	.66
	12	.54	.57
	13	.56	.52
	14	.59	.65
	15	.64	.72

Relationships Between *IARA* Scores and Measures of Mathematics Achievement

Barron, Ansley, and Hoover (1991) collected a sample of 825 9th grade students from Algebra I classes in three Midwestern school districts. These students took the *IARA* early in the fall. Their grades for both the first and second semesters were recorded as were their scores from the *Iowa Tests of Basic Skills (ITBS,* Hieronymus & Hoover, 1986) from grade eight and their scores from the *Iowa Tests of Educational Development (ITED,* Feldt, Forsyth, & Alnot, 1989) from the fall of grade nine. It should be noted that the scores on the *IARA* were not used to place these students into Algebra I classes. It should also be noted that this group of students was above average relative to the *IARA* norm group. The mean composite score for this group was approximately 159, which implies that the typical student in this group fell roughly at the 73rd national percentile (1991 norms).

The intercorrelations of the *IARA* composite scores, the two semester exams, and the semester grades are given in Table A8.

It is clear from these values that the *IARA* composite score is highly related to both Algebra I grades and test scores. These correlations provide some evidence that the *IARA* composite score could be useful in deciding whether a student should be placed in Algebra I. Validity evidence for the *IARA* is also provided by its relationship to other tests assessing similar content. The *IARA* composite scores were significantly related to both the Mathematics Total scores of the *ITBS* (r=.69) and the Quantitative Thinking scores of the *ITED* (r=.48).

To augment this validity evidence, multiple regression analyses were carried out using the *IARA* composite scores and the *ITBS* Mathematics Total scores as predictors of Algebra I test scores and grades. Of particular interest was whether the *IARA* scores could add significantly to the accurate predictions of the four criterion variables, given that the *ITBS* Mathematics Total scores were already available. The rationale for this was that most schools probably have available scores from standardized tests. If the *IARA* scores cannot significantly contribute to the accuracy of prediction of success in Algebra I above and beyond the information provided by data already on hand, the usefulness of the *IARA* would be in doubt. From these regression analyses, it was found that the *IARA* composite scores did indeed significantly add to the prediction of success in Algebra I.

Barron, Ansley, and Hoover (1991) also constructed expectancy tables for this sample. Two examples of these expectancy tables are given in Tables A9 and A10. To use this information in choosing students who are ready to enter Algebra I, consider the following example. Suppose a student obtains a composite score of 145. Only about 25% (6.0% + 18.9%) of the students in the validity study at approximately this score level earned first semester marks above C. It is important to note that the typical student in this validity study scored at the 73rd percentile. If a school has a higher average than that of the validation group, these expectancy figures most likely underestimate the percentages of successful students. On the other hand, if a school's average is less than that of the validation group, there is a strong likelihood that these values overestimate the percentages of successful students. These tables can be used for reference purposes until a school has gathered enough of its own data to create a local expectancy table. The construction of such a table would be an invaluable addition to the local decision-making process. However, it must be stressed again that the *IARA* scores should only augment and never replace the professional judgments of teachers. It is also important to note that this validation study was done with the 4th edition of the *IARA*. The national norms referenced there were from 1991. A standard score of 145 from the 4th edition roughly corresponds to the same national percentile rank as a standard score of 145 on the 5th edition (PR=37).

Table A8
Correlations Between the *IARA* Composite Score and Algebra Grades and Test Scores

Variable	2.	3.	4.	5.
1. *IARA* Composite	.69(.84*)	.49(.75*)	.65(.82*)	.45(.74*)
2. 1st Semester Exam		.59	.76	.56
3. 1st Semester Grade			.58	.78
4. 2nd Semester Exam				.58
5. 2nd Semester Grade				

*Correlations have been corrected for restriction in range.

Table A9
Percent of Students Earning the Various Possible First Semester Algebra I Grades for *IARA* Composite Score Intervals

IARA Composite	First Semester Algebra Grades					
SS Intervals	A	B	C	D	F	N
101–139	4.7	6.3	51.6	37.5	0.0	64
140–149	6.0	18.9	54.5	20.2	0.8	134
150–159	10.5	30.9	42.4	15.2	1.1	191
160–169	23.8	38.1	31.4	6.4	0.4	252
170–179	42.1	37.9	16.8	2.1	1.1	95
180–199	56.2	29.2	14.6	0.0	0.0	89
N	181	246	295	98	5	825

Table A10
Percent of Students Earning the Various Possible Second Semester Algebra I Grades for *IARA* Composite Score Intervals

IARA Composite	Second Semester Algebra Grades					
SS Intervals	A	B	C	D	F	N
101–139	6.3	9.4	32.8	40.6	10.9	64
140–149	6.0	20.2	41.8	27.6	4.5	134
150–159	8.9	28.3	34.6	24.6	3.7	191
160–169	28.6	29.0	28.6	10.3	3.6	252
170–179	34.7	43.2	16.8	5.3	0.0	95
180–199	52.8	28.1	12.4	6.7	0.0	89
N	181	226	242	147	29	825

References

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: Author.

Barron, S., Ansley, T., and Hoover, H. (1991). *Gender differences in predicting success in high school algebra.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

Common Core State Standards. © Copyright 2010. National Governors Association Center for Best Practices and Council of Chief State School Officers. All rights reserved

Dorans, N and Holland, P. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P.W. Holland and H. Wainer, (Eds.), *Differential item functioning* (pp. 35-66). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Ebel, R. and Frisbie, D. (1991). *Essentials of educational measurement* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall, Inc.

Feldt, L., Forsyth, R., and Alnot, S. (1989). *Iowa tests of educational development.* Iowa City: The University of Iowa.

Hieronymus, A. and Hoover, H. (1986). *Iowa tests of basic skills.* Iowa City: The University of Iowa.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author.

National Education Database™. Quality Education Data. (2004). Denver, CO: Author

Witt, E., Ankenmann, R., and Dunbar, S. (1996). *The sensitivity of the Mantel-Haenszel statistic to variations in sampling procedures in DIF analysis.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, New York.