The Natural World Test, Version 9 (NW-9)

A measure of quantitative and scientific reasoning

Test Manual

THE CENTER FOR ASSESSMENT & RESEARCH STUDIES

MSC 6806
Harrisonburg, VA 22807
540.568.6706 Phone
540.568.7878 Fax
assessment@jmu.edu
www.jmu.edu/assessment

Donna L. Sundre
2008

Donna L. Sundre
Amy Thelk & Clifton Wigtil
October, 2008

Table of Contents

The Natural World, Version 9 (NW-9)

The NW-9 Test Manual

Section 1. The Nature of the Instrument

The Natural World Test, Version 9 (NW-9) is a 66-item multiple choice test developed by science and mathematics university faculty to assess college students' quantitative and scientific reasoning skills. The Natural World Test is currently in its ninth edition and reflects development spanning over the last decade.

Section 2. Intended Use

2.1. Appropriate and inappropriate uses and interpretations

Currently, the NW-9 is intended to provide information about the effects of curriculum and instruction on students' learning. Therefore, test results may be used to inform curriculum and instructional improvements at the program or institution level. However, it is not appropriate to use test results to make decisions about individual students. Currently, psychometric properties are not sufficient to support high-stakes classifications for individuals (please refer to section 5.2 *Evidence of Reliability*). This instrument was also not intended as a vehicle for providing individual students with feedback about their mastery of quantitative reasoning. Institutions may choose to provide their students with individual feedback, but results should *not* be used to make any type of high-stakes classification decisions. According to the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2000), test users are responsible for collecting validity evidence for any uses of the test other than those recommended here.

2.2. Target population

The NW-9 is intended to measure learning in scientific and quantitative reasoning for undergraduate college students. Since items were designed to be content-free, this instrument should be appropriate for students in any general education science curriculum. Although the NW-9 was designed with all undergraduate college students in mind, it is important to take note of the normative sample when using this instrument. Please refer to Section 5.4 *Norming* for more information.

2.3. Qualifications of users

Test users must be trained to administer assessments in standardized conditions. The *Proctor qualifications and training section* of this manual (Section 4.1) provides more information about how proctors can be trained for test administration. James Madison University will complete scoring[1] for this instrument. In addition, test users should be knowledgeable about how to interpret the statistical results from the test and how to make appropriate inferences about the program using the results. Test users who do not have a measurement background or do not have in-depth knowledge of the program are encouraged to consult with colleagues who have the necessary knowledge.

Section 3. Test Development

3.1. Academic and theoretical basis

The NW-9 was designed to evaluate student learning in eight general education objectives related to scientific and quantitative reasoning.

- Objective 1: Describe the methods of inquiry that lead to mathematical truth and scientific knowledge and be able to distinguish science from pseudo-science
- Objective 2: Use theories and models as unifying principles that help us understand natural phenomena and make predictions
- Objective 3: Recognize the interdependence of applied research, basic research, and technology, and how they affect society
- Objective 4: Illustrate the interdependence between developments in science and social and ethical issues
- Objective 5: Use graphical, symbolic, and numerical methods to analyze, organize, and interpret natural phenomena

[1] Scoring (correct/incorrect) includes all instrument items and, if desired, up to 20 additional items added to the instrument by those institutions seeking to further evaluate the construct.

- Objective 6: Discriminate between association and causation, and identify the types of evidence used to establish causation
- Objective 7: Formulate hypotheses, identify relevant variables, and design experiments to test hypotheses
- Objective 8: Evaluate the credibility, use, and misuse of scientific and mathematical information in scientific developments and public-policy issues

3.2. Item type selection

All NW-9 items use a selected-response format, with the number of response options ranging from two to six. The items were written as such to ease scoring, to maintain objective scoring, and to minimize test-taker fatigue. Most items follow a typical multiple-choice format, in which an item stem is followed by alternative responses consisting of the correct answer and at least one distracter. An effort was made for items on the NW-9 to have only three response options, one correct answer and two high-quality distracters. Items with three or more distracters were often part of a set of items in which the same alternatives were provided for each (similar to a matching item design).

3.3. Item pool and scale development process

Open-ended interviews with mathematics and science faculty members provided information about the construct and objectives. Guided by this information, items were written by faculty in direct relation to the objectives. A back translation (Dawis, 1987) was subsequently conducted to determine how items matched back to the appropriate objective. This activity requires panelists to consider each item individually and to attempt to identify which objective(s), if any, the item addresses.

Starting with the eighth version of the Natural World test, faculty also completed a content alignment activity to establish that the test items did correspond with the objectives which they were intended to assess. In contrast to the back translation, the content alignment exercises ask each panelist to consider each objective, one at a time, then to go through the exam to locate items which might contribute to the measurement of the objective. The content alignment method is preferable to the back translation since judges tend to assign an item to a single objective during the back translation procedure, even if the item is a good match to more than one objective (Miller, Setzer, Sundre & Zeng, 2006).

Once the back translation and content alignment activities were completed, the items were administered to a small pilot study group to determine how easy the instrument's instructions were to follow and to examine test format, length of test, and the appropriateness of the items to the college student population. Next, the test was administered to a random sample of first-year college students (with no previous college experience) and then to a randomly selected sample of students in the spring semester of their second year (with varying amounts of experience with their general education courses).

The original version of the test has since undergone multiple revisions based on item and content analyses. For example, items from the NW-6 that performed well statistically (*i.e.*, had consistently high item-total correlations over multiple administrations) were used on the NW-7. Items from previous versions of the NW that were discarded were revised and included on NW-7. Items submitted by faculty members in previous years that were not used in any previous version of the NW were also revised and added to the NW-7. More recently, an item-writing workshop for faculty members, held in 2004, resulted in the development of several of the NW-7 items. An item-analysis of the NW-7 provided evidence that some items did not perform well. These items were removed, subsequently forming the NW-8 version of the test. Finally, during the summer of 2007, faculty from various disciplines in science and math convened and reviewed the test. Items that were construed as problematic were revised, and additional items were developed to address gaps in content coverage, leading to the current version of the test, the NW-9.

3.4. Subscores and their development

The test blueprint for the NW-9 appears in Table 1. Some items are mapped to more than one objective; thus, the number of items assessing each objective sums to a value greater than the total number of items on the test.

The first administration of the NW-9 was performed in fall 2007. The results show that internal consistency value for the overall test is reasonably high, with a reliability (coefficient alpha) value of .78. However, use of subscale scores is not recommended at this time, due to lower than desirable levels of internal consistency (i.e., less than .50). These scores and other information on the test can be reviewed in the Technical Information section to follow.

Table 1

Test Blueprint for NW-9

NW-9 Objectives	# Items	Items
1. Describe the methods of inquiry that lead to mathematical truth and scientific knowledge and be able to distinguish science from pseudo-science.	13	2, 5, 9, 14, 18, 28, 38-41, 55-57 (19.7% of total test)
2. Use theories and models as unifying principles that help us understand natural phenomena and make predictions.	7	17, 20, 22, 27, 64-66 (10.6% of test)
3. Recognize the interdependence of applied research, basic research, and technology, and how they affect society.	7	1, 15, 16, 43-46 (10.6% of total test)
4. Illustrate the interdependence between developments in science and social and ethical issues.	9	2, 19, 24-26, 29, 55-57 (13.6% of total test)
5. Use graphical, symbolic, and numerical methods to analyze, organize, and interpret natural phenomenon.	21	4, 7, 8, 10-13, 21, 30-33, 51-53, 58-63 (31.8% of total test)
6. Discriminate between association and causation, and identify the types of evidence used to establish causation	10	3, 34-37, 53, 60-63 (15..2% of total test)
7. Formulate hypotheses, identify relevant variables, and design experiments to test hypotheses.	21	5, 6, 9-13, 18, 23, 28, 41, 42, 47-50, 54, 59, 60, 62, 63 (31.8% of total test)
8. Evaluate the credibility, use, and misuse of scientific and mathematical information in scientific developments and public-policy issues.	13	2, 14, 24-26, 29, 38-40, 60-63 (19.7% of test)
Quantitative Reasoning	26	3, 4, 7, 8, 10-13, 21, 30-37, 51-53, 58-63
Scientific Reasoning	49	1, 2, 5, 6, 9-20, 22-29, 38-50, 54-57, 59-66
Total Test	66	1-66

Section 4. Administrative Procedures

4.1. Proctor qualifications and training

While administration of the NW-9 does not require intense training, proctors should be given guidance on standardized test administration. Proctor training can be accomplished in a brief session in which they are familiarized with the test instructions and the general procedures to be adhered to during the test administration. During training, proctors should be provided with the standardized instructions to be used in the actual testing session. Instructions for each mode of administration are provided in the following section.

4.2. Testing procedures

The NW-9 can be administered in either paper-and-pencil or computer-based formats. Administration procedures for each method vary slightly and are presented separately below. Regardless of administration method, examinees should be provided with scrap paper and a pencil. Additionally, room temperature and lighting should be appropriate for optimal testing performance. Before beginning the test, examinees should be provided with general information about the number, type and content of items on the test. Examinees should be informed of the amount of time they will be given to complete the test and what they should do upon completion of the test. It is recommended that students be given at least 60 minutes to complete the NW-9. However, if the testing time is 60 minutes and the majority of students are still working after 55 minutes, the proctor may decide to extend the testing period for another 5 minutes. When the testing time is almost over (for example, after 55 minutes), the proctor should periodically announce the time remaining (e.g. 5 minutes, 2 minutes, 1 minute).

4.2.1. Paper and pencil administration

If paper and pencil administration of the NW-9 test is elected, examinees receive an item booklet and an answer sheet such as a Scantron form. If a Scantron form is used as an answer sheet, a form that can accommodate up to six response options is required. For example, the General Purpose Form (number 6703) by Pearson NCS has been used successfully for the NW-9. Examinees should be given instructions on how to identify themselves for the test (if necessary) and how to record answers on the answer sheet properly.

Below are the instructions that should be read out loud by the proctor to the students prior to beginning the test. These are the instructions used during the testing sessions which generated the data for use in this manual. *If you elect to give your test with different instructions than those presented here, please understand that you will be administering the test under non-standardized conditions. As a result, the scores you obtain may be incomparable to the normative data offered in this manual.*

The following standardized instructions are to be read immediately preceding administration of the paper-and-pencil NW-9 test.

> "You are about to take a test which will help us understand the thinking processes you use to understand the world. You will be given 60 minutes to respond to 66 multiple choice questions. Mark all of your answers on the answer sheet; use the scrap paper provided to work on any problems; please do not write on the test item booklet**.** If you have a question during the testing period or need additional scrap paper, raise your hand and a proctor will come to you. An announcement will be made when there are 5 minutes remaining. If you finish early, please turn over your answer sheet and remain quiet until the conclusion of the test."

4.2.2. Computer-based administration

If computer administration of the NW-9 test is elected, students will each be seated in front a computer on which the NW-9 test has been set up in Adaptex testing software. Students should receive either oral or written instructions on how to enter their identification information and begin the test. These standardized instructions are read immediately preceding administration of the computer-administered NW-test, and all data generated for this manual were collected in conditions which utilized these instructions. The instructions which follow should also be provided on the computer screen just prior to administering the first item.

> "You are about to take a test which will help us understand the thinking processes you use to understand the world. You will be given 60 minutes to respond to 66 multiple choice questions. If you have a question during the testing period, or need additional scrap paper, raise your hand and a proctor will come to you. An announcement will be made when there are 5 minutes remaining. If you finish early, please turn over your instruction sheet and scrap paper, turn off your monitor, and remain quiet until the conclusion of the test."

Please see the Appendix for information about troubleshooting for issues which may arise during computer-based testing.

4.3. Extent of exchangeability

When an instrument is administered in both paper-and-pencil and computer-based formats, information derived from one context may not be directly applicable to another context. In other words, it cannot be assumed that reliability or validity information collected through one mode of administration will generalize to the other mode (Mead & Drasgow, 1993; Wise & Plake, 1990). Equivalence between testing settings should be established before applying information collected in one setting to information collected in another.

Very small differences in scores were found to exist between paper-based and computer-based administrations, though the effect size was shown to be small. In general, students completing the computer-based NW-9 had no significant differences with the pencil-and-paper version. There are several factors which may influence the effect of the administration method. For example, computer-administered testing could be problematic for test takers who are unfamiliar with using a computer. In such situations, computer skills may confound the meaning of NW-9 test scores. To avoid this, it is recommended that either a trained proctor be available to assist with computer skills, detailed instructions on basic computing skills be provided, or basic computer competency be a prerequisite of taking the computer based test. Also, even computer savvy test-takers may encounter technical problems during a testing session that cannot be anticipated or prevented (i.e. computer freezing or power shortage). If a situation like this occurs, test scores may be unusable or biased due to

frustrations or interruptions. It is recommended that a computer technician oversee the computers being used for testing and be available during testing for troubleshooting.

Besides technical problems, there are other factors to consider when selecting an administration method. For example, students taking the paper-based test could skip items (with the possibility of returning to them later), whereas students taking the computer-based test may not be able to skip items (they may be forced to make a selection to move on), or they may not be able to review previously answered items depending on how the computer-based assessment is set up. These factors may influence the comparability of scores across administration methods. Therefore, these set-up options should be carefully considered and discussed with a local computer-based testing administrator.

One advantage of the computer-based administration method is that test-takers cannot be unclear about their response or provide invalid (out-of-range) responses. This results in less missing data. However, in the NW-9 test, this may influence the comparability of scores obtained from the different administration methods because missing data is scored the same as incorrect data (zero).

As shown in Table 2, in general, students completing the computer-based NW-9 performed as well as those completing the paper-based form. The differences between the two administrations in NW-9 scores were not statistically significant. sizes (*d*, a standardized value, can be interpreted in terms of number of standard deviations).

Table 2

Mean, Standard Deviations and Reliabilities of NW-9 Scores Obtained in Computer-Based and Paper-Based Administrations

	Computer-based				Paper-based				Differences			
Sample	***N***	***M***	***SD***	**α**	***N***	***M***	***SD***	**α**	***t***	**df**	***p***	***d***
Freshmen Fall 2007	413	41.9	6.8	.73	995	42.3	7.7	.79	0.84	1406	0.399	0.05
Sophomores Spring 2008	97	47.2	5.9	.65	923	46.9	8.0	.81	-0.308	1018	0.758	-0.03

Section 5. Technical Information

5.1. Scoring and interpretation

All NW-9 items are multiple-choice. When creating a total score for an examinee, correct responses to items are given a score of '1' and incorrect responses to items are given a score of '0.' The correct responses are then summed. A high total score would suggest that an examinee has high levels of quantitative and scientific reasoning, and a low total score would suggest than an examinee has low levels of quantitative and scientific reasoning. Similarly, the Quantitative Reasoning subscore can be obtained by summing the number of items that are answered correctly to that subscale (i.e., items which assess objectives 5 and 6). Summing the scores on the items assigned to the Scientific Reasoning subscale (i.e. items assessing objectives 1-4, 7, and 8) results in the Scientific Reasoning subscore.

5.2. Evidence of Reliability

Reliability refers to the degree of stability and consistency of test scores. Due to the various sources of variability in test scores, there are different ways of measuring reliability. The NW-9 has been examined for reliability as measured by Cronbach's alpha (α), which is frequently used to determine internal consistency. Specifically, α requires only one administration and is the mathematical equivalent of the average of all possible split-half reliability computations. Alpha indicates how much variance in the observed scores is attributable to the true score. In other words, α indicates how related the scores on the items are to the construct of interest (in this case, quantitative and scientific reasoning). While coefficients with a value of .70 or higher have traditionally been considered adequate for scale use, reliabilities above .80 are desirable (Nunally, 1978).

The instrument has been given in two administrations: Fall 2007 and Spring 2008. The corresponding reliabilities (expressed using Cronbach's α) of the overall test are presented in Table 3.

Table 3

Sample Sizes, Cronbach's Alpha, Raw Scores, and Standard Deviations for NW-9

Sample	*N*	α	*M*	*SD*
Fall 2007	1,408	.79	43.19	7.70
Spring 2008	1,020	.80	46.96	7.83

5.3. Evidence of Validity

Validity refers to the degree to which one can make inferences from the scores obtained on a test. Validity is not an absolute state, but rather a collection of evidence indicating that the scores obtained on a test are valid for their intended use (AERA, 2000).

Faculty from mathematics and science departments wrote the NW-9 items using the objectives as their guide, and conducted a content alignment of the NW-9 items to further verify the extent to which items were linked to the appropriate objectives. During this most recent mapping, every item in the NW-9 successfully translated back to an objective.

Since the NW-9 is a measure of achievement of general education in quantitative and scientific reasoning, it is realistic to expect students with more course experience to outperform students with fewer course credits in those areas. Freshmen may enter the university with Advanced Placement (AP) credits or transfer credits for other institutions; these courses may fulfill the general education requirement for scientific and quantitative reasoning. Therefore, their exposure to these courses should result in higher scores as compared to students with no college-level experience in these domains. For advanced students (sophomores, or those with 45-70 completed credits), students have related courses experience either from credits brought in with them as freshmen, or through completion of general education courses in quantitative and scientific reasoning. In tables 4 and 5, differences in scores based on related course experience is presented.

Table 4

Freshmen Scores by Number of College-Level Math and Science Courses (or the Equivalent) Taken

# of related science/math courses	N	Raw score Mean	Raw score SD
0	1173	41.45	7.32
1	162	45.62	7.16
2	45	46.58	7.49
3	16	44.81	6.87
4 or more	12	46.92	5.99

Table 5

Sophomore Scores by Number of College-Level Math and Science Courses (or the Equivalent) Taken

# of related science/math credits	N	Raw score Mean	Raw score SD
0	30	43.53	8.95
1	245	46.23	7.85
2	332	47.08	7.52
3	231	47.43	7.29
4 or more	132	48.42	9.13

5.4. Norming

At the time of this writing, the NW-9 had been administered to one random sample of students at JMU. Data was collected in fall 2007 and consisted of a sample of 1,408 incoming first-year students. The test was administered in a low-stakes environment. Under these testing conditions, students may not give their best effort; therefore, the scores may underestimate student knowledge of the intended construct. The ethnic backgrounds of the students in the sample roughly approximated those of the overall JMU population (83% white, 5% Asian, 4% Black, 2.5% Hispanic, and 5% not specified, and 0.2% other).

To determine how the students at your institution performed in relation to college students at the institution that serves as the site of NW-9 research, refer to Tables 6 and 7. Table 6 contains score information for first-year students. Table 7 displays score information for the sophomores (students with 45-70 completed credits). The percentile ranks associated with each NW-9 raw score are presented for the total group and by gender.

Table 6

*Percentile Ranks for NW-9 Scores for **Freshmen** at a Mid-Atlantic 4-Year Institution*

Score	Total Group (*N* = 1408*)	Females (*n* = 872)	Males (*n* = 530)
63	99.93	--	99.81
62	--	--	--
61	99.82	--	99.53
60	99.68	99.94	99.25
59	99.47	99.77	98.96
58	99.18	99.60	98.49
57	98.83	99.43	97.83
56	98.12	99.03	96.60
55	97.05	98.22	95.09
54	95.63	96.90	93.49
53	93.86	95.24	91.51
52	91.55	93.12	88.87
51	88.74	90.65	85.47
50	85.65	88.36	81.04
49	81.36	85.03	75.09
48	76.92	80.96	70.00
47	72.87	76.95	65.85
46	67.93	72.25	60.47
45	62.61	67.26	54.53
44	57.39	62.16	49.06

43	52.45	57.11	44.34
42	46.91	51.26	39.34
41	41.30	45.18	34.43
40	36.54	39.85	30.57
39	31.75	34.52	26.79
38	27.24	29.76	22.83
37	23.37	25.86	18.96
36	19.96	22.19	15.94
35	16.76	18.46	13.58
34	13.96	15.19	11.51
33	11.51	12.27	9.81
32	9.30	9.40	8.77
31	7.71	7.51	7.74
30	6.43	6.14	6.60
29	4.90	4.47	5.28
28	3.44	2.81	4.15
27	2.73	1.89	3.87
26	2.31	1.61	3.30
25	1.81	1.15	2.74
24	1.56	0.86	2.55
23	1.38	0.63	--
22	1.03	0.29	2.08
21	0.60	--	1.32
20	0.39	--	0.85
19	0.25	--	0.47
18	0.11	0.06	--
17	0.04	--	0.09
16 or fewer	--	--	--

-- indicates that this score was not present in the data set.
* Gender data missing for 6 individuals; these students did receive scores on the NW-9.

Table 7

*Percentile Ranks for NW-9 Scores for **Sophomores** at a Mid-Atlantic 4-Year Institution*

Score	Total Group (*N* = 1020*)	Females (*n* = 647)	Males (*n* = 372)
64	99.90	--	99.73
63	99.51	99.85	98.92
62	99.07	99.61	98.12
61	98.53	99.30	97.18
60	97.70	98.84	95.70
59	96.23	97.68	93.68
58	94.36	96.21	91.13
57	92.50	95.05	88.04
56	89.80	92.97	84.27
55	86.08	89.72	79.70
54	81.27	85.24	74.33
53	77.16	81.53	69.49

52	73.38	77.67	65.86
51	67.65	71.48	60.89
50	61.67	65.53	54.84
49	57.01	60.82	50.27
48	51.96	55.41	45.83
47	46.32	49.15	41.26
46	41.52	44.13	36.83
45	36.81	38.95	32.93
44	31.52	32.77	29.17
43	26.32	26.89	25.13
42	22.35	22.41	22.04
41	19.31	19.01	19.62
40	16.18	15.46	17.20
39	13.82	12.91	15.19
38	12.25	11.36	13.58
37	10.49	9.66	11.69
36	8.73	7.96	9.95
35	7.01	6.18	8.47
34	5.59	4.64	7.26
33	4.61	3.55	6.45
32	3.97	2.78	6.05
31	3.38	2.24	5.38
30	2.99	--	4.70
29	2.75	1.85	4.30
28	2.50	1.62	--
27	--	--	--
26	2.21	1.31	3.76
25	1.81	1.00	3.23
24	1.47	--	2.42
23	1.23	0.85	--
22	0.98	0.54	1.75
21	0.69	0.23	1.48
20	0.44	0.08	1.08
19	0.20	--	0.54
18	0.05	--	0.13
17 or fewer	--	--	--

-- indicates that this score was not present in the data set.
* Gender data missing for 1 individual; this student did receive a score on the NW-9.

5.5. Meeting a standard

5.5.1. The Standard Setting Process

In 2008, 37 science and math faculty members participated in an Angoff standard setting (Angoff, 1971). This process involves having participants, or panelists, give judgments about the likelihood that a "minimally competent" student could get the item correct. For this purpose, the minimally competent student is someone who completed all of their quantitative and scientific reasoning general education requirements by just passing the courses (i.e., no interest in the area beyond filling requirements, and receiving average grades in these classes). Panelists make a judgment for each item, expressed in terms of a probability carried out two decimal places (e.g. 50% probability that student would be able to complete the item correctly would be recorded as .50). The ratings for each item are then summed over the entire test; the resulting total is the

panelist's cut score for the exam. After each panelist's score was computed, the median of the all the cut scores was set as the cut sore.

5.5.2. Faculty expectations

Students who have science and/or math credits when they arrive at JMU should be closer to meeting the competency or academic standard than those with no general education credits. Similarly, sophomores with more coursework should do better than their peers with fewer completed courses in related areas.

Table 8 shows the faculty determined cut score and the scores achieved by students, as well as proportion of students in each category who met standards.

Table 8

Fall 2007 First-Year Student and Spring 2008 Sophomore/Junior Performance Compared with Faculty Expectation

	Faculty Standard		Freshmen with no related coursework *n*=1173	All sophomores *n*=973	Sophomores with no related coursework *n*=10	Sophomores who completed general education requirements *n*=156
NW-9 Total (66 items)	50.4	Mean	42.4	47.0	45.7	48.3
		Percent meeting standard	*13.5%*	*36.5%*	*30.0%*	*47.4%*

Section 6. Additional Information

Additional information on the NW-9 may be obtained by contacting the Center for Assessment and Research Studies, James Madison University, MSC 6806, Harrisonburg, VA 22807. Information may also be obtained through the following Web site:
http://www.jmu.edu/assessment/resources/Overview.htm

Section 7. References

American Educational Research Association, American Psychological Association, & National Council of Measurement in Education. (2000). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Angoff, W. A. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement, 2nd edition* (pp. 508-600). Washington, DC: American Council on Education.

Dawis, R. (1987). Scale construction. *Journal of Counseling Psychology, 34*(4), 481-489.

Mead, A. D. & Drasgow, F. (1993). Equivalence of computerized tests and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin, 114,* 449-458.

Miller, B. J., Setzer, C., Sundre, D. L., & Zeng, X. (2007, April). Content validity: A comparison of two methods. Paper presentation to the National Council on Measurement in Education. Chicago, IL

Nunally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Wise, S. L. & Plake, B. S. (1990). Computer-based testing in higher education. *Measurement and Evaluation in Counseling and Development, 23,* 3-10.

Section 8. Appendix

Troubleshooting guide for issues arising in computer-based test administration

For computer-based administration, students take the NW-9 using Adaptex, a web-based software program developed at JMU. Students respond to test items by using the computer mouse to select the correct option using radio buttons. When using this method, test administrators should be prepared to troubleshoot any technical difficulties that may occur during testing. The instructions below provide steps to take if technical difficulties should arise.

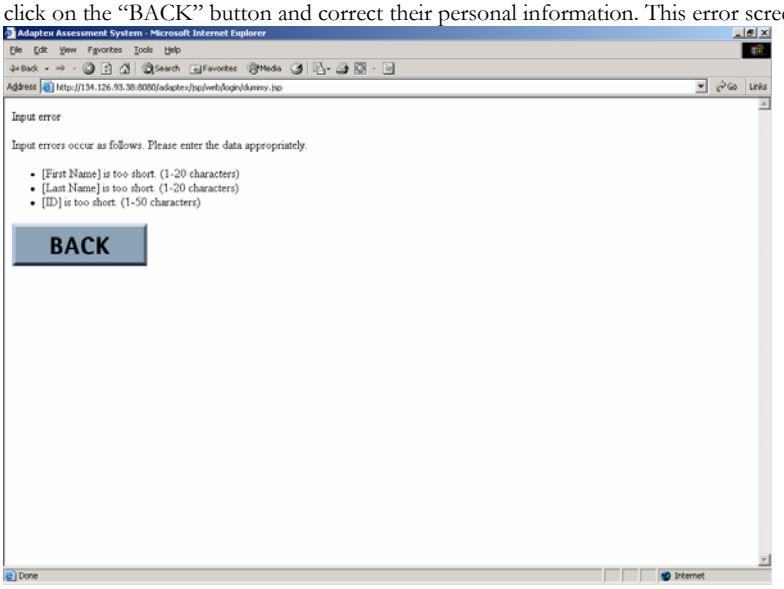
Troubleshooting in Computer-Based Administration

To Recover a Test:
Close the browser if it is still open.
Open the Internet Explorer browser and type in the URL that the student was given. Click **CONTINUE**.

1. Fill in the student name and ID in the appropriate boxes and select the institution on the login screen.
2. Select the appropriate test and type the password. Click **SUBMIT**.
3. You will be prompted whether or not you want to resume the test you previously did not finish. Click **CONTINUE**. The program will take you to the question the student was completing when the problem arose.

Page Cannot Be Displayed Error:
If students get "Page cannot be displayed" errors, they can click the "Back" link to go back to the last question on which they were working.

Input Error Message:
This message appears when students have entered personal information incorrectly. If this screen appears, students should click on the "BACK" button and correct their personal information. This error screen appears as below:



Notes for proctors

Students should not run any programs before or during the test. As the students arrive, please ask them to take a seat at a computer but DO NOT let them play on the computers. (When students open music files or play on the computers the files they access take up storage in the cache memory, which creates a problem for the program to run efficiently, so NO ONE is allowed to play at any time on the computers.)

There are two web addresses provided. Please make sure there are approximately equal numbers of students who receive each URL because the server cannot handle too many students.

Restart the computers between test sessions to clear out the computer memory.

If the program crashes or encounters any problems, you may recover the test (See Troubleshooting section in manual)

Make sure that the screen resolution is set to 1024x768.

Students should not double-click the Next/Back button, as it may cause a skipped question.