8-1-2006

The Effects of Item Grouping on Test Reliability

Holly Mackey
Western Kentucky University

Follow this and additional works at: http://digitalcommons.wku.edu/theses

Part of the Psychology Commons

Recommended Citation
http://digitalcommons.wku.edu/theses/294

This Thesis is brought to you for free and open access by TopSCHOLAR®. It has been accepted for inclusion in Masters Theses & Specialist Projects by an authorized administrator of TopSCHOLAR®. For more information, please contact topscholar@wku.edu.
THE EFFECTS OF ITEM GROUPING ON TEST RELIABILITY

A Thesis
Presented to
The Faculty of the Department of Psychology
Western Kentucky University
Bowling Green, Kentucky

In Partial Fulfillment
Of the Requirements for the Degree
Master of Arts

By
Holly Casebier Mackey

August 2006
THE EFFECTS OF ITEM GROUPING ON TEST RELIABILITY

Date Recommended  June 16, 2006

Director of Thesis

[Signature]

Dean Graduate Studies and Research  Date

[Signature]  6/21/06
# Table of Contents

Abstract........................................................................................................... iv

Introduction...................................................................................................... 1

Method........................................................................................................... 9

Results........................................................................................................... 11

Discussion...................................................................................................... 14

References..................................................................................................... 17

Appendix A: Test Script.................................................................................. 19

Appendix B: Grouped Test.............................................................................. 22

Appendix C: Ungrouped Test.......................................................................... 29

Appendix D: Test Instructions.......................................................................... 38
Optimal ordering of items on a multidimensional test has been the focus of several studies. In all but one study, previous research centered on measures of personality or opinion. The current study examines item grouping effects for a cognitive ability test. Two forms of a cognitive ability test containing four constructs (verbal ability, basic computation, number series, and spatial visualization) were prepared. Form A consisted of items grouped by construct, and Form B had items dispersed randomly throughout the test. The order of items within a construct remained the same for both forms. Tests were administered to 186 undergraduate psychology students. Coefficient alpha estimates of reliability on Form A were compared to coefficient alphas for the same constructs on Form B. Additionally, differences in mean scores across all four constructs and construct intercorrelations were compared by format. There were no significant differences in coefficient alphas and only one (basic computation) was in the hypothesized direction. There was only one significant construct intercorrelation pairing (basic computation vs. spatial visualization), and there were no significant differences in mean scores. Based on the lack of consistent findings, we found little support for a grouping effect for cognitive ability tests.
Introduction

Instrument development can prove challenging in any context, but it is particularly challenging when the instrument being developed is multidimensional; that is, a single instrument which attempts to measure multiple constructs. Not only is it necessary to ensure that the items measure the intended construct, but there is the additional decision of whether all items should be dispersed throughout the instrument at random or if items measuring the same construct should be grouped together. In the past, some researchers decided to ignore the possibility that item grouping could potentially affect internal consistency, reliability, scale intercorrelation, and discriminant and convergent validity; they distributed items throughout their instruments without concern for the impact of their placement (Schriesheim & DeNisi, 1980).

Despite the emergence of research in this area (Hunt, 2005; Mailloux, 2002; Melnick, 1993), there remains no clear answer as to whether the grouping of items has any effect. The results of previous research have been mixed. As such, there have been arguments for both sides of the issue. Most previous research has centered on measures of personality or attitude with few studies dealing with performance based measures.

Arguments for Grouping Items Together

Several arguments have been made to support grouping items by construct. Item grouping may serve to relieve fatigue experienced by respondents, which is not only a common source of error in measurement but also a common complaint among test takers. Additionally, grouping items can reduce the monotony of lengthy instruments (Schriesheim & DeNisi, 1980) and improve readability (Metzner & Mann, 1953), which may improve the favorability of the instrument among respondents.
Schriesheim and DeNisi (1980) discussed an annoyance that many respondents share when completing a personality inventory. Respondents often believe that they are being asked the same question multiple times in an attempt by the researcher to catch them being dishonest. Grouping together all items measuring the same construct allows respondents to see that although some questions are similar, they are not being asked to answer identical questions multiple times. Thus, grouping can help to reduce mistrust of the researchers on the part of the respondents.

Finally, grouping items allows for continuity of thought which may enhance the quality of performance (Schriesheim & DeNisi, 1980). When items are grouped, respondents are not forced to repeatedly switch their response set back and forth between constructs (referred to in the literature as *switching gears* and *task set switching*). Task set switching is a cognitive mechanism in which individuals switch between cognitive processes depending on the particular tasks at hand. A task set switch requires cognitive control in situations for which a particular stimulus could be processed in a variety of different ways. In these instances, the cognitive system has the burden of organizing itself correctly according to the instructions for the given task (Altman, in press). The response time necessary to switch between tasks is the time cost, and the associated difference in error rates is the error cost. These costs associated with switching between tasks provide important clues about the mechanisms at work when task set switching is necessary. The time it takes the cognitive system to prepare for the new stimulus determines how long it will take an individual to produce a response to a new task. When presented with a new task, the cognitive system must reconfigure itself with the appropriate set of stimulus-response rules in order to process the information required to complete the task (Yeung &
Monsell, 2003). When items are grouped by construct, respondents are better able to focus their attention on all of the items which measure a single construct independent of the influence of the other constructs. This reduced burden on the cognitive system has the potential to enhance the quality of responses (Melnick, 1993).

*Arguments for Randomly Distributing all Items*

Schriesheim (1981) argued that randomizing items throughout the instrument might serve to reduce leniency errors that could otherwise be committed by respondents. He argued that grouping items together might make it easier for respondents to base their responses on evaluative biases which could result in increased rating leniency. Leniency bias refers to instances where a particular rater consistently rates everyone above average rather than providing an accurate evaluative rating for each individual being rated. Metzner and Mann (1953) stated that when items are grouped in a questionnaire, the relationships that are found may be, to some extent, imposed by the designer of the instrument. Melnick (1993) argued that when items are grouped, previous items could contaminate future responses. That is, respondents could use cues from previous items to answer future questions. Although it is possible for this influence to happen when items are not grouped, the grouping of the items could increase the likelihood that item contamination could occur.

A final argument for randomizing items throughout the instrument is that there are times, particularly in personality testing, in which the intent of the test or the identity of the construct measured needs to remain covert in order to obtain an honest response from test takers. When items are grouped based on the construct, it becomes much easier for respondents to identify the construct that is being measured. With this knowledge of the
construct, respondents could alter their responses in an attempt to provide the answers they believe are desired (Schriesheim & DeNisi, 1980).

**Previous Research on Item Grouping**

Research seeking to examine the effect of item grouping has been in short supply and the results of these studies have been inconsistent. As such, it is difficult to ascertain whether item grouping actually enhances the quality of a test. The first researchers to try and answer that question were Metzner and Mann (1953). They developed two forms of a questionnaire, one form in which items measuring the same construct were grouped together and labeled and another form which placed items throughout the form without regard to the construct measured. These forms were distributed to respondents at random, and the relationships between questions measuring the same constructs were explored. Metzner and Mann found some support for the hypothesized grouping effect; however, a few issues with their study should be mentioned. First, they analyzed only 13 items from the questionnaire rather than the entire instrument. These 13 items measured four factors, with two of the factors based on only two items each. Also, they analyzed correlations between adjacent items rather than examining all items concurrently. As a result, the reader is given no indication as to how the test was affected as a whole.

Nearly 30 years after Metzner and Mann (1953), Schriesheim and DeNisi (1980) and then Schriesheim (1981) were the next researchers to explore item grouping effects. Both of these studies employed the same design as the Metzner and Mann study and used a combination of two leader behavior questionnaires as their measure. Schriesheim and DeNisi argued that grouping the items together may actually make it easier for respondents to make quick decisions about a dimension without giving careful thought to
each item within the dimension. Although Schriesheim and DeNisi hypothesized that convergent and discriminant validity would improve in the grouped condition, the results of their study did not support the hypothesis. Schriesheim (1981) was concerned with how leniency bias was influenced by the grouping of items on an instrument. In his study, only internal consistency indices were examined, and again, the results did not support a grouping effect. It is important to note, however, that the total sample size was only 60 (30 people per condition) in both Schriesheim and DeNisi and Schriesheim. Unfortunately, a sample size of this magnitude reduces the statistical power to a level that renders the results of both of the studies uninterpretable.

Schurr and Henriksen (1983) performed a study which used a 61-item measure of attitudes of teaching behavior using Metzner and Mann’s (1953) research design. The results of their study were mixed. Just two of the six analyses revealed significant grouping effects. Additionally, Schurr and Henriksen failed to offer any explanation as to why these results were found.

Allison (1984) conducted the only study which did not use a personality or attitude measure. He used a sixth grade science test and distributed it to over 300 students. On the grouped version of the test, items relating to a given topic (animal classification, solar system, and earth science) were grouped together. In the second version of the test, all items were dispersed throughout the exam at random. As with many of the item grouping studies, Allison failed to find any internal consistency differences.

Melnick (1993) administered a 41-item questionnaire which measured educational administrator attitudes of six facets of teacher behavior (3 to 9 items per factor). The
results of this study indicated that there may be differences in internal consistency between the grouped and random versions of the questionnaire, but because no significance tests were performed on these differences, the results are not interpretable.

The first study to report clear findings in favor of a grouping effect was conducted by Mailloux (2002). He developed a questionnaire consisting of seven dimensions (3 to 6 items per factor) involving issues related to attending college. Mailloux collected a sample of over 1,200 high school students, and like others, he constructed a grouped and random version of the questionnaire. Unlike others, his results showed not only a significant increase in internal consistency for the grouped version of the questionnaire, but these differences were also found across all race based comparisons. The biggest improvement in reliability among the two versions of the forms was with found with samples of Hispanic and African American students.

Finally, Hunt (2005) conducted a study examining the effects of item grouping on the internal consistency and construct intercorrelation of a personality test measuring extraversion and conscientiousness. No significant differences were found between the internal consistencies of the grouped versus random version of the personality questionnaire for each of the constructs. Additionally, no significant differences were found between scale intercorrelations for the two versions of the questionnaire. It should be noted that in the randomized version of the questionnaire, the questions were alternated between constructs such that the first question measured extraversion and the second conscientiousness. This simple pattern may have been easy for respondents to identify. This awareness of the pattern of constructs in the randomized version could have contributed to the nonsignificant results. Additionally, only two constructs were used,
which may not be enough to produce a grouping effect. That is, a test measuring only two constructs may not force respondents to change their response set. Similarly, a limited change would not be sufficiently cognitively demanding to produce grouping effects.

It is clear that the results of the research in this area are very inconsistent. Given the heavy use of psychological measures in mental health, employment selection, and education, item grouping is an area that needs additional investigation to ensure that the format of the instruments is not influencing the responses of the test taker.

*The Present Study*

The current study will further investigate the effects of item grouping on the internal consistency of tests. This study will use an optimal performance test designed to measure four constructs: verbal analogies, basic mathematical computations, number series, and spatial visualization. The performance based method was chosen for the current study for several reasons. First, the Allison (1983) study is the only study that has used this type of measure. Second, the cognitive demand of a performance based test is such that respondents are forced to search for a correct answer to the question rather than simply provide a response for which there is no right or wrong answer (e.g., attitude or personality measurement). The previously discussed issues associated with continuity of thought and task set switching may not present themselves when the task is of low cognitive demand, as is the case with measures of personality and attitude. If the cognitive system does not have to restructure itself in order to provide responses to items of low cognitive demand, measures of personality and attitude would not be expected to produce a grouping effect. The present study seeks to provide suggestions for the development of performance based tests which are based on empirical research. The
implications of a grouping effect would be far reaching. If it is possible to develop tests which measure the intended construct with more accuracy, both test developers and test takers would benefit. The Hypotheses of the present study are as follows:

Hypothesis 1: Coefficient alpha estimates of reliability will be greater when items are grouped by construct than when the items are ungrouped (i.e., random construct order).

Hypothesis 2: Construct intercorrelations will be lower when items are grouped by construct than when items are ungrouped.

Hypothesis 3: Test scores will be greater when items are grouped by construct than when items are ungrouped.
Method

Participants

Undergraduate psychology students at a large, public, southeastern university served as participants in this study. In partial fulfillment of their course requirements, Introduction to Psychology students were required to participate in a number of research studies. Students chose among several research studies which were available to them through the university’s study board. In addition to those students who volunteered as a means of fulfilling their course requirements, other students volunteered to participate in this study as one possible means of gaining extra credit in other psychology courses. Students who decided not to participate in the current study had a variety of alternative ways to gain extra credit.

Materials

Two versions of the test were developed, a grouped version and an ungrouped version. Both versions consisted of the same 40 items accompanied by the same set of instructions (see Appendix A for the grouped test, Appendix B for the ungrouped test, and Appendix C for the test instructions). On the grouped version of the test, items were arranged according to construct, such that all ten verbal items were followed by the ten basic computation items, the ten number series items, and finally the ten spatial visualization items. The items within each construct were placed in a random order with no concern for item difficulty or any other facet of the item.

On the ungrouped version of the test, the 40 items were dispersed throughout the test at random without regard to construct; however, the items within each construct appeared in the same order throughout each version of the test. For example, the sixth
verbal item that was presented on the grouped version of the test was also the sixth verbal item that was presented on the random version of the test. This consistent ordering ensured that the only differences between test versions was whether the items were grouped by construct, eliminating the possibility that item order within construct could confound the results.

Randomization was achieved by placing all items in a hat and drawing them out to establish the order in which they would appear on the test. Other published tests, such as the Adaptability Test developed by Tiffin and Lawshe (1942), have used this same random order format.

Procedure

Test booklets, instructions for the test, and informed consent forms were prepared and inserted into manila folders to be distributed to the participants. Before the participants arrived at the designated research location, the two versions of the test were randomly distributed among the chairs in the room, such that half of the participants would receive the grouped version of the test and the other half would receive the ungrouped version. This method of distribution ensured that each participant had an equal opportunity to receive either version of the test.

Prior to their participation, participants were given information regarding the nature of the study before they were asked to decide whether they wanted to consent to participate in the research. The instructions were then read to each group of students from a prepared script to ensure that everyone received the same instructions (see Appendix D for script).
Results

Data were collected from 186 undergraduate psychology students. Among the participants, 149 (80%) were Caucasian, 18 (10%) were African American, 9 (4.7%) stated that none of the race categories described them, 2 (1%) were Hispanic, and 8 (4.3%) failed to report a race. There were 75 male participants (40%) and 106 female participants (57%). Five participants (3%) did not report their sex. The ages of the participants ranged from 17 to 52 ($M = 19.85, SD = 2.97$). Each testing condition (grouped and ungrouped) contained 93 participants. Responses from participants were scored by hand using a scoring rubric. While scoring items, it was determined that one of the number series items was written incorrectly. Thus this item (the last number series item on each version of the test) was excluded from the analysis. After all tests were scored, item responses were entered into a computer database by hand for subsequent analysis.

The first hypothesis stated that coefficient alpha estimates of reliability would be greater when items were grouped by construct than when items were distributed throughout the test in an ungrouped fashion. These data are displayed in Table 1. None of the differences between coefficient alphas were significant, and only one (basic computation) was in the hypothesized direction. Due to the nonsignificant results for all four constructs, it was concluded that Hypothesis 1 was not supported.

The second hypothesis stated that construct intercorrelations would be lower when items were grouped by construct as compared to the ungrouped ordering. First, total scores on verbal, number series, basic computation, and spatial visualization items were computed for each participant. These total scores were correlated with each other.
Table 1

*Coefficient Alphas for Each Construct by Condition*

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Grouped</th>
<th>Ungrouped</th>
<th>z statistic for difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal</td>
<td>.21</td>
<td>.31</td>
<td>-.72</td>
</tr>
<tr>
<td>Number Series</td>
<td>.35</td>
<td>.41</td>
<td>-.47</td>
</tr>
<tr>
<td>Basic Comp.</td>
<td>.73</td>
<td>.68</td>
<td>.66</td>
</tr>
<tr>
<td>Spatial</td>
<td>.82</td>
<td>.84</td>
<td>.37</td>
</tr>
</tbody>
</table>

*Note.* $N = 93$ for each group. $p > .05$. for all comparisons, one-tailed test.

Ideally, correlations between constructs should be low (i.e., better discriminant validity). These data are presented in Table 2. Construct intercorrelations were compared by condition (grouped vs. ungrouped) for all six possible construct pairs. Only one pairing yielded a significant difference (basic computation vs. spatial visualization). Due to the lack of consistent statistical differences, it was concluded that Hypothesis 2 was not supported.

The final hypothesis stated that mean scores would be greater when items were grouped by construct than when items were distributed throughout the test in an ungrouped fashion. Scores were calculated for each of the four constructs for both the grouped and ungrouped version of the test (eight mean scores total).

Independent samples $t$-tests were used to test the differences between group means for significance for each of the four constructs. These results are displayed in Table 3. There were no significant differences between group means. Thus, Hypothesis 3 is not supported.
Table 2

*Construct Intercorrelations*

<table>
<thead>
<tr>
<th>Construct Pairs</th>
<th>Grouped r</th>
<th>Ungrouped r</th>
<th>z test for difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal vs. Basic Comp.</td>
<td>.07</td>
<td>.19</td>
<td>0.81</td>
</tr>
<tr>
<td>Verbal vs. Number Series</td>
<td>.08</td>
<td>.27</td>
<td>1.29</td>
</tr>
<tr>
<td>Verbal vs. Spatial</td>
<td>.26</td>
<td>.41</td>
<td>1.15</td>
</tr>
<tr>
<td>Basic Comp. vs. Number Series</td>
<td>.38</td>
<td>.29</td>
<td>-0.74</td>
</tr>
<tr>
<td>Basic Comp. vs. Spatial</td>
<td>.17</td>
<td>.19</td>
<td>0.17</td>
</tr>
<tr>
<td>Number Series vs. Spatial</td>
<td>.16</td>
<td>.49</td>
<td>*2.57</td>
</tr>
</tbody>
</table>

*Note. N = 93 for each group. *p < .05, one-tailed test.*

Table 3

*Mean Test Scores for each Construct by Dimension*

<table>
<thead>
<tr>
<th>Construct</th>
<th>Grouped</th>
<th>Ungrouped</th>
<th>t statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Verbal</td>
<td>4.74</td>
<td>1.61</td>
<td>5.23</td>
</tr>
<tr>
<td>Number Series</td>
<td>6.78</td>
<td>2.66</td>
<td>6.49</td>
</tr>
<tr>
<td>Basic Comp.</td>
<td>5.85</td>
<td>2.56</td>
<td>6.66</td>
</tr>
<tr>
<td>Spatial</td>
<td>6.64</td>
<td>1.94</td>
<td>6.72</td>
</tr>
</tbody>
</table>

*Note. N = 93 for each group. p > .05 for all comparisons, one-tailed test.*
Discussion

As with prior research on item grouping, the hypotheses of this study were not supported. Scales were not more internally consistent when the items were grouped by construct than when the items were ungrouped. With one exception, scale intercorrelations were not lower when the items were grouped by construct as compared to ungrouped items. Finally, grouped item mean scores were not significantly greater than ungrouped item mean scores.

These results fail to support the hypothesis that item grouping on multidimensional cognitive abilities tests enhance the quality of the test. Based on the current study, it appears that tests with ungrouped test items work as well as tests with grouped test items. However, there are some limitations to the present study which may have contributed to the lack of support. First, the types of items that were chosen for the two versions of the test (verbal, basic math computation, number series, and spatial visualization) may not have required participants to develop strategies for solving the problems. If participants were not forced to switch their thought processes (task set switching) as they switched between constructs, a grouping effect would not occur.

Another limitation of the current study is the length of the instruments used. There were 40 items on each version of the test. Due to the cognitive demand of the test items and the length of the test, participants (who had little external incentive to answer items correctly) may have simply worked quickly to complete the items without exerting any additional effort to answer the items correctly. Related to the overall length of the instrument is the number of items per construct. There were only ten items per construct which may not be enough to produce a task set. More items may be necessary in order to
engage any task switching. An instrument with more items per construct and external motivators for participants might be able to produce a grouping effect.

A problem that became apparent while scoring the test items was that the last number series item on both versions of the test (Item 20 on the grouped test, Item 39 on the ungrouped test) was written incorrectly. As a result, this item was thrown out before any analysis was done. One can only speculate as to whether this had any impact on the participants as they worked to complete the items. It could be that participants became frustrated when they could not solve this item and then put less effort into solving subsequent items. Also, participants could have spent an excessive amount of time trying to solve this item and then rushed to complete the other items.

Others have suggested reasons why an ungrouped test format might be superior to the grouped version. For example, Schriesheim and DeNisi (1980) speculated that dispersing items throughout the test might reduce the monotony associated with answering multiple questions which assess the same construct. It is possible that individuals who are not as competent with one construct as compared to the other constructs, may find it intimidating when ten such items are presented to them en masse (i.e., as with the grouped version of the test). When the challenging items are spread out among the other test items, some of the anxiety the test taker experiences when answering the challenging item type might be reduced. The increased anxiety in the grouped format could inhibit the test taker’s performance.

Prior research has focused heavily on personality testing and opinion surveys. The present study is only the third such study with a focus on cognitive abilities tests. Future research should focus on this exploration of item grouping for tests of cognitive ability.
As mentioned, certain groups of tasks require a great deal of cognitive control in order to switch among tasks. Future research should explore which types of tasks contribute to the largest amount of cognitive control in task switching and then develop test items based on those tasks to see if a grouping effect is generated. It is possible that only a limited set of test items function differently in a grouped format.

Finally, it may be that tests work equally well in both a grouped and ungrouped format. It may be that participants simply prefer tests that are grouped, but they are able to perform the same regardless of the format. Future research should explore participant perception of test format. If both grouped and ungrouped are equal, but participants prefer the grouped format, test developers would be wise to chose a grouped version to increase test taker satisfaction. This issue would be particularly relevant in preemployment testing where companies benefit from using the most applicant friendly methods of selection.
References


Appendix A:

Test Script
Script

Hello, my name is ___________. Today, you are being asked to participate in a psychological research study. This study should take approximately 30 minutes. You will be asked to take a test which consists of 40 questions. Please do your very best on this test. Before going any further, I need everyone who is planning to participate in the study to fill out an informed consent form. This form provides additional information about the study and must be signed in order for you to participate. Please notice that your participation is voluntary and you may withdraw from the study at anytime without penalty. I need you to write in your student ID number on the informed consent form so that you will be able to receive credit for participation in this study.

(PASS OUT TEST FOLDER)

When the test folder is handed to you, you will notice that there are two sets of stapled materials inside. Only pull out the informed consent packet. Please leave the test booklet inside the folder. Please read through the informed consent form, decide if you would like to participate, and then if you decide you would like to participate; sign the form. Place your Student ID number next to your name on the consent form, so that you will be able to receive credit for your participation. This consent form will be kept separate from your responses at all times to ensure that your responses are anonymous.

You will notice that there are instructions for the test stapled to the back of the informed consent form. Please tear the instruction form off of the consent form before passing it in. We will be going over these instructions together once all forms have been signed and turned in. (allow time to read consent form; then ask if everyone is finished; pick up forms).
Now, we will go over these instructions together, and I will answer any questions you might have before you begin the test. Please do not take the test booklet out of the folder until told to do so.

(Collect Informed consent forms)

You may now take out your test booklet. At the top of the first page, there is a space for you to fill in information indicating your age, race, and sex. Also notice that test booklets are printed front and back and that there are 4 pages in each booklet. There are 40 items total. Everyone should check, at this time, to make sure they have a complete test booklet. What questions do you have at this time?

(answer questions)

If anyone has any question throughout the test, please raise your hand and someone will come to assist you. When you have completed all of the test items, place the test booklet back into the folder, and someone will come and take it from you. When everyone has finished, I will give you a little bit of information about the study. You may now turn the page and begin answering the questions.

(when everyone has finished and all tests have been gathered)

I would like to thank each of you for your participation in this study. Some of you were given a test which had items grouped together by construct. This means that all of the items measuring your verbal ability were grouped together, followed by all of the items measuring your basic math computation ability, and so on. Others were given a test form in which all items were dispersed throughout the test at random with no regard to which construct they were measuring. The purpose of this study is to determine in which instance the test works better. It is hypothesized that the test where items were grouped by construct will have better internal consistency than the version which had the items dispersed randomly. I would like to take this time to thank everyone again for your participation in this study.
Appendix B:

Grouped Test
Please write-in or circle your... age _____

sex: male or female

race: Hispanic, African-American, White, or none of the above

1. FRICTION : ABRASION
   A. sterility : cleanliness
   B. dam : flood
   C. laceration : wound
   D. heat : evaporation
   E. literacy : ignorance

2. OVERDOSE : PRESCRIPTION
   A. deprivation : materialism
   B. indiscretion : convention
   C. affliction : sympathy
   D. adventure : expedition
   E. drug : medicine

3. EVAPORATE : VAPOR
   A. petrify : stone
   B. centrifuge : liquid
   C. saturate : fluid
   D. corrode : acid
   E. incinerate : fire

4. SHARD : POTTERY
   A. flint : stone
   B. flange : wheel
   C. cinder : coal
   D. fragment : bone
   E. tare : grain

5. MERCENARY : MONEY
   A. vindictive : revenge
   B. scholarly : library
   C. immaculate : cleanliness
   D. thirsty : water
   E. belligerent : invasion

6. DOLPHIN : MAMMAL
   A. larva : insect
   B. penguin : bird
   C. sonnet : stanza
   D. computer : machine
   E. peninsula : island

7. DRAWL : SPEAK
   A. spurt : expel
   B. foster : develop
   C. scintillate : flash
   D. pare : trim
   E. saunter : walk

8. MANSION : RESIDENCE
   A. limousine : automobile
   B. chandelier : candle
   C. tuxedo : wardrobe
   D. diamond : rhinestone
   E. yacht : harbor
9. **ENVELOPE : LETTER**
   A. scarf : hat
   B. box : bag
   C. crate : produce
   D. neck : head
   E. blood : heart

14. 20, 38, ______, 74.
   A. 48
   B. 12
   C. 32
   D. 56
   E. 70

10. **CHOREOGRAPHY : DANCE**
   A. ceremony : sermon
   B. agenda : advertisement
   C. poetry : recitation
   D. instrumentation : conductor
   E. plot : story

15. 5, 25, ______, 390625.
   A. 50
   B. 625
   C. 225
   D. 2050
   E. 500

16. 9, 16, ______, 36.
   A. 32
   B. 25
   C. 18
   D. 20
   E. 23

17. 2, 5, 11, 23, 47, ______.
   A. 73
   B. 94
   C. 95
   D. 104
   E. 72

18. ______, 15, 5, -5.
   A. 5
   B. 10
   C. 15
19.  6, _______, -2, -6.
   A. 2  D. 8
   B. 4  E. 0
   C. 6

20.  8, 27, 64, _______, 125, 216.
   A. 12  D. 52
   B. 27  E. 48
   C. 36

21. A. 6 x 23 = 148
    B. 7 x 85 = 595
    C. 3 x 59 = 178
    D. 6 x 56 = 316
    E. 3 x 82 = 236

22. A. 606/3 = 192
    B. 480/8 = 80
    C. 392/4 = 88
    D. 228/19 = 12
    E. 2336/73 = 42

23. A. $12(120) = 10$
    B. $(12)(2)(120) = 8$
    C. $12(24) = 2$
    D. $(60)120 = 300$
    E. $(12)(120) = 12$

24. A. 9 x 32 = 268
    B. 7 x 63 = 441
    C. 4 x 73 = 282
    D. 14 x 17 = 248
    E. 8 x 19 = 172

25. A. 2(7x3) = 52
    B. -8(6x9) = -432
    C. -9(-4x-3) = 108
    D. -7(3x2) = 32
    E. 12(3x8) = 278
26.  
A. $\frac{888}{3} = 296$  
B. $\frac{777}{3} = 279$  
C. $\frac{222}{3} = 64$  
D. $\frac{555}{3} = 175$  
E. $\frac{444}{3} = 128$

27.  
A. $2^2 \times 14 = 66$  
B. $4^2 \times 12 = 198$  
C. $5^3 \times 5 = 525$  
D. $3^3 \times 9 = 243$  
E. $6^3 \times 2 = 422$

28.  
A. $17.5 \times 20 = 275.5$  
B. $9.5 \times 32 = 314$  
C. $8.75 \times 13 = 123.75$  
D. $13.2 \times 16 = 211$  
E. $21.25 \times 7 = 148.75$

29.  
A. $223 + 553 = 766$  
B. $65 + 1235 = 1310$  
C. $432 + 978 = 1420$  
D. $834 + 235 = 1069$  
E. $78 + 452 = 550$

30.  
A. $13 \times 23 = 299$  
B. $15 \times 18 = 260$  
C. $17 \times 34 = 588$  
D. $21 \times 54 = 1124$  
E. $41 \times 12 = 482$

31. [images of five different patterns]
Appendix C:

Ungrouped Test
Please write-in or circle your...

age ____ sex: male or female

race: Hispanic, African-American, White, or none of the above

1. A. 6 x 23 = 148   D. 6 x 56 = 316
   B. 7 x 85 = 595   E. 3 x 82 = 236
   C. 3 x 59 = 158

2. 

3. A. 606/3 = 192   D. 228/19 = 12
   B. 480/8 = 80  E. 2336/73 = 42
   C. 392/4 = 88

4. 

5. FRICTION : ABRASION
   A. sterility : cleanliness
   B. dam : flood
   C. laceration : wound
   D. heat : evaporation
   E. literacy : ignorance
7. OVERDOSE : PRESCRIPTION
   A. deprivation : materialism
   B. indiscretion : convention
   C. affliction : sympathy
   D. adventure : expedition
   E. drug: medicine

8. EVAPORATE : VAPOR
   A. petrify : stone
   B. centrifuge : liquid
   C. saturate : fluid
   D. corrode : acid
   E. incinerate : fire

9. A. \( \frac{12(120)}{120} = 10 \)
    B. \( \frac{(12)(2)(120)}{180} = 8 \)
    C. \( \frac{12(24)}{180} = 2 \)
    D. \( \frac{(60)120}{24} = 300 \)
    E. \( \frac{(12)(120)}{180(2)} = 12 \)

10. 1, 1, 2, 3, 5, 8, 13, ______.
    A. 26  
    B. 13  
    C. 39  
    D. 21  
    E. 31
11. SHARD : POTTERY
   A. flint : stone
   B. flange : wheel
   C. cinder : coal
   D. fragment : bone
   E. tare : grain

12. A. 9 x 32 = 268                        D. 14 x 17 = 248
    B. 7 x 63 = 441                        E. 8 x 19 = 172
    C. 4 x 73 = 282

    A. 2073                               D. 4703
    B. 3003                               E. 3703
    C. 3773

14. MERCENARY : MONEY
   A. vindictive : revenge
   B. scholarly : library
   C. immaculate : cleanliness
   D. thirsty : water
   E. belligerent : invasion

15. -17, ______, 27, 49.
    A. 5                                  D. 10
    B. -7                                 E. 21
    C. 7
16.  
A. \(2(7 \times 3) = 52\)  
B. \(-8(6 \times 9) = -432\)  
C. \(-9(-4 \times -3) = 108\)  
D. \(-7(3 \times -2) = 32\)  
E. \(12(3 \times 8) = 278\)  

17.  
20, 38, _____, 74.  
A. 48  
B. 12  
C. 32  
D. 56  
E. 70  

18.  
A. \(888/3 = 296\)  
B. \(777/3 = 279\)  
C. \(222/3 = 64\)  
D. \(555/3 = 175\)  
E. \(444/3 = 128\)  

19.  

20.  
5, 25, _____, 390625.  
A. 50  
B. 625  
C. 225  
D. 2050  
E. 500  

21.  
A. \(2^2 \times 14 = 66\)  
B. \(4^2 \times 12 = 198\)  
C. \(5^3 \times 5 = 525\)  
D. \(3^3 \times 9 = 243\)  
E. \(6^3 \times 2 = 422\)
22. 9, 16, ______, 36.
   A. 32  D. 20
   B. 25  E. 23
   C. 18

23. A. 17.5 x 20 = 275.5
    B. 9.5 x 32 = 314
    C. 8.75 x 13 = 123.75
    D. 13.2 x 16 = 211
    E. 21.25 x 7 = 148.75

25. 2, 5, 11, 23, 47, ______.
   A. 73  D. 104
   B. 94  E. 72
   C. 95

26. DOLPHIN : MAMMAL
   A. larva : insect
   B. penguin : bird
   C. sonnet : stanza
   D. computer : machine
   E. peninsula : island
28. DRAWL : SPEAK
A. spurt : expel
B. foster : develop
C. scintillate : flash
D. pare : trim
E. saunter : walk

29. ______, 15, 5, -5.
A. 5 
B. 10 
C. 15 
D. 20 
E. 25

30. 6, ______, -2, -6.
A. 2 
B. 4 
C. 6 
D. 8 
E. 0

31. 
A. $223 + 553 = 766$
B. $65 + 1235 = 1310$
C. $432 + 978 = 1420$
D. $834 + 235 = 1069$
E. $78 + 452 = 550$

32. 

33. 

35. MANSION : RESIDENCE
   A. limousine : automobile
   B. chandelier : candle
   C. tuxedo : wardrobe
   D. diamond : rhinestone
   E. yacht : harbor

36. ENVELOPE : LETTER
   A. scarf : hat
   B. box : bag
   C. crate : produce
   D. neck : head
   E. blood : heart

37. A. $13 \times 23 = 299$
    B. $15 \times 18 = 260$
    C. $17 \times 34 = 588$
    D. $21 \times 54 = 1124$
    E. $41 \times 12 = 482$

38. CHOREOGRAPHY : DANCE
   A. ceremony : sermon
   B. agenda : advertisement
   C. poetry : recitation
   D. instrumentation : conductor
   E. plot : story
39. 8, 27, 64, ______, 125, 216.
A. 12    D. 52
B. 27    E. 48
C. 36

40. [diagram with four options]
Appendix D:

Test Instructions
Do your best on this test. Work as rapidly as you can without making unnecessary mistakes. When you find that you can not answer a question, do not spend an excessive amount of time on it, simply move on to the next question. **DO NOT** go back to answer questions that you left blank, and do not skip around to other questions, rather complete the items in the order in which they appear on the test.

**SAMPLE QUESTION 1**
Which of the word pairs in the answer choices has the same type of relationship as the given word pair?

- ADULT : CHILD
- A. horse : mare
- B. cat : kitten
- C. swine : sow
- D. human : animal
- E. cow : herd

First, try to decide what each of the words mean. You should focus on how the word ADULT relates to the word CHILD. Next circle the letter of the answer choice in which the two words relate to each other in the same manner as ADULT relates to CHILD. In this example, a child grows up to be an adult, and a kitten grows up to be a cat. **The Correct Answer is B.**

**SAMPLE QUESTION 2**

2, 0, 4, 2, 6, 4, 8, ___, 10

- A. 4
- B. 6
- C. 2
- D. 0
- E. 3

You should figure out what is being done in each number series in order to determine what number completes the series. In this example, 2 is being subtracted, then 4 is being added each time. **The number that completes the series is 6, thus you should circle B.**

**SAMPLE QUESTION 3**

A: 6 x 12 = 74  
B: 23 x 6 = 148  
C: 4 x 13 = 62  
D: 14 x 19 = 266  
E: 7 x 14 = 108

You should check each calculation and circle the letter of the item that shows the correct answer to the given problem. **You should circle the letter D, because 14 x 19 = 266.**

**SAMPLE QUESTION 4**
Determine which two of the four drawings on the right show the same object as the one on the left. There are always two correct answers for each problem. **Put an X under the two correct drawings.**