Western Kentucky University TopSCHOLAR®

Masters Theses & Specialist Projects

Graduate School

5-1-2007

An Examination of the Effects of Item Grouping on Internal Consistency and Scale Intercorrelation

Julian Hunt Western Kentucky University

Follow this and additional works at: http://digitalcommons.wku.edu/theses Part of the <u>Psychology Commons</u>

Recommended Citation

Hunt, Julian, "An Examination of the Effects of Item Grouping on Internal Consistency and Scale Intercorrelation" (2007). *Masters Theses & Specialist Projects*. Paper 500. http://digitalcommons.wku.edu/theses/500

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 connie.foster@wku.edu.

AN EXAMINATION OF THE EFFECTS OF ITEM GROUPING ON INTERNAL CONSISTENCY AND SCALE INTERCORRELATION

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 Julian T. Hunt

May 7, 2005

AN EXAMINATION OF THE EFFECTS OF ITEM GROUPING ON INTERNAL CONSISTENCY AND SCALE INTERCORRELATION

Date Recommended Apr. 1 20, 2005

Director of Thesis

thom? WOO.

05 me

Dean, Graduate Studies and Research Date

1 abic of Coments	Table	of	Contents
-------------------	-------	----	----------

Abstractiv
Introduction1
Method
Results
Discussion
References 12
Appendix A: Informed Consent
Appendix B: Experimental Script 15
Appendix C: Grouped Questionnaire
Appendix D: Ungrouped Questionnaire

AN EXAMINATION OF THE EFFECTS OF ITEM GROUPING ON INTERNAL CONSISTENCY AND SCALE INTERCORRELATION

Julian T. HuntMay 7, 200522 PagesDirected by: Reagan Brown, Ph.D., Betsy Shoenfelt, Ph.D., Anthony Paquin, Ph.D.Department of PsychologyWestern Kentucky University

Developers of multidimensional tests must decide whether to group items measuring the same construct together or list the items randomly. The effects of item grouping are not well established by the few existing studies. This study examined the effects of item grouping on the psychometric properties of a personality inventory measuring conscientiousness and extraversion. Two hundred and four undergraduate students were administered the test with the items listed in either a grouped or ungrouped format. Internal consistency reliability was estimated with coefficient alpha. Grouping test items failed to increase scale reliability (.80 for ungrouped versus .76 for grouped) or decrease scale intercorrelation (.31 for grouped items versus .27 for ungrouped items). The differences between these correlations were not significant. Reasons for the findings are offered and recommendations are given for future studies.

Introduction

Professionally developed tests are carefully constructed to ensure that they measure only the desired constructs. Tests that measure a single construct (i.e., unidimensional) consist of a homogeneous item pool. Tests that measure more than one construct (i.e., multidimensional) consist of a mixture of items measuring each relevant construct. Researchers have commonly distributed items throughout a test without regard to the effects of item order (Schriesheim & DeNisi, 1980). With a multidimensional test, researchers can list first all the items that measure Factor 1 followed by all the items that measure Factor 2 (i.e., grouped items). Conversely the researcher can randomize or alternate the order of items between factors (i.e., ungrouped). Item grouping potentially impacts internal consistency reliability, scale intercorrelation, and discriminant and convergent validity.

Item Grouping Arguments

Reasons given for and against item grouping are varied. Schriesheim and DeNisi (1980) noted that optimal performance test developers believed items should be labeled and grouped according to dimensions. Grouping and labeling by construct should allow for continuity of thought and enhance the quality of performance. Additionally, grouping items may break the monotony of the questionnaire and reduce the effect of fatigue, a source of error in measurement (Schriesheim, 1981). Respondents can become bored with questionnaires that are extreme in length. One reason for boredom among respondents during lengthy questionnaires is the similarity of item style. By grouping the items that measure the same constructs, respondents may realize that the questions are not duplicates (Schriesheim). Finally, grouping items does not force test takers to shift their

1

response set after each question (Melnick, 1993); that is, test takers can focus on a single construct until all of the items measuring that construct have been completed, potentially enhancing the quality of the responses.

There are two arguments against item grouping. Melnick (1993) stated that when items are grouped by construct, prior items could influence test takers in a manner that could contaminate their subsequent responses. Although the same problem could occur with ungrouped items, the consistency of content in a grouped test could make the contamination more likely to occur. A second argument against grouping is specific to measures of personality and attitudes. In this case grouping could facilitate test taker awareness of the construct measured. This awareness could then help test takers intentionally distort their responses to achieve a desired goal (Schrieshiem & DeNisi, 1980).

Previous Research on Item Grouping

The number of studies investigating item grouping is limited, and as such, it is difficult to conclude whether item grouping has any effect on the psychometric properties of tests. The first such study was by Metzner and Mann (1953). The design of their study (subjects randomly assigned to one of two forms of the same questionnaire) has been replicated many times since. Metzner and Mann found some, but not clear, support for a grouping effect. Two issues with their study should be considered. First, they analyzed only 13 items from a larger pool of items. These 13 items measured four factors, with one factor based on only two items. Additionally, their analysis focused on the correlation between adjacent items rather than a simultaneous consideration of all items for a given

dimension (or even total scores per factor), offering no indication as to how the test as a whole was affected (e.g., dimensionality).

The next item grouping studies were by Schriesheim and DeNisi (1980) and Schriesheim (1981). Both studies used total scale scores in their analysis. Schriesheim and DeNisi investigated item grouping effects on internal consistency and construct validity, whereas Schriesheim examined only internal consistency reliability. Unfortunately, the total sample size in each study was 60 (30 per group), a sample so lacking in statistical power that it renders their results uninterpretable.

In a study not plagued by sample size limitations (total N = 721), Schurr and Henriksen (1983) performed a factor analysis of a 61-item measure of attitudes of teaching behaviors. Their results showed grouping effects for two of the six sets of analyses. Schurr and Henriksen were unable to offer any explanations as to why grouping effects were found on just the two sets of analyses. Allison (1984) conducted the only examination of grouping effects not focused on a measure of personality or attitudes. Allison examined item grouping effects on an optimal performance science test for sixth grade students and failed to find significant internal consistency differences. Finally, Melnick (1993) administered a 37-item questionnaire measuring six factors of workplace issues (3 to 9 items per factor). Although Melnick examined internal consistency reliability, he failed to test the differences between the coefficients for significance.

There appears to be little consistency in item grouping research, and thus, it is difficult to draw conclusions. Metzner and Mann (1953) probably used too few items to find an effect. Both Schriesheim and DeNisi (1980) and Schriesheim (1981) suffered from sample sizes far too small for research involving correlations. Melnick (1993) failed

to test the difference between correlations for significance. Allison (1984) appears to have few if any conceptual problems but is also the only study to examine grouping effects with an optimal performance test. Given the differences between the constructs, it is entirely possible that item grouping may have an effect only for behavioral, attitudes, or personality measures. Finally, Schurr and Henriksen (1983) found some grouping effects but were unable to explain why their results were not consistent for all their analyses. In summary, additional item grouping research appears warranted.

The Present Study

The purpose of this study is to further examine the effect item grouping has on the internal consistency and scale intercorrelation of a personality test. I plan to explore how item grouping affects test taker responses on a questionnaire measuring two dimensions of normal personality: extraversion and conscientiousness.

Grouped items, as opposed to ungrouped items, allow test takers to consider a single construct at a time. Focusing on the same construct for multiple items in a row should increase the consistency of their responses to items of that construct (Melnick, 1993). Additionally, grouped items should help clarify the difference between constructs; that is, after answering a number of very similar items, the presence of a different item should be clear to a test taker. This newfound awareness that the concept behind the questions has changed should assist the test taker in providing a response that is different from the responses given to previous items.

Hypothesis 1: A questionnaire with grouped items will have higher internal consistency reliability coefficients than will a questionnaire with ungrouped items.

Hypothesis 2: A questionnaire with ungrouped items will have a higher scale intercorrelation (i.e., the correlation between the scores for each construct) than will a questionnaire with grouped items.

Method

Participants and Procedure

A total of 204 undergraduate college students participated in this study. A signed informed consent (see Appendix A for an example) was obtained from all students before starting the experiment. The participants were randomly assigned to either a grouped or ungrouped condition. All data were collected anonymously. A copy of an experimenter's script for the conduct and procedure of the experiment can be found in Appendix B. *Materials*

In the questionnaire for the grouped condition the items were listed separately by construct without construct names at the beginning of each item set. The names of the constructs were omitted to conform with standard questionnaire construction practice. Although a listing of the construct names might enhance the grouping effect and has been done in some group studies (e.g., Schriesheim & DeNisi, 1980), it is a departure from typical practice. The ungrouped questionnaire contained the same items as the grouped one, but the order of the items was alternated between constructs. An alternating order is similar to the item order on the NEO-FFI (Costa & McRae, 1992), which presents items measuring five constructs in an alternating order. As a means to separate item grouping effects from item order effects, the order of the items relative to the other items of the same construct remained the same as in the grouped questionnaire.

The two constructs measured are extraversion and conscientiousness. The items utilized are the extraversion and conscientiousness items from the NEO-FFI. The NEO-FFI is a short form of the NEO-PI-R (Costa & McRae, 1992), a professionally developed measure of the five factor model of personality. The shorter NEO-FFI was chosen

6

because the length of the NEO-PI-R may discourage participants from completing the questionnaire thoughtfully. All of the participants were undergraduate students, and their main motivation for completing the questionnaire was the extra credit they received for their class. A copy of the grouped item questionnaire can be found in Appendix C. A copy of the ungrouped questionnaire is located in Appendix D.

Analysis

Internal Consistency reliability was estimated with Cronbach's coefficient alpha. Differences between coefficient alphas (as well as differences between scale intercorrelations) were tested for significance with the *z*-test for differences between correlations from independent samples.

Results

Data were collected from 204 undergraduate students. Of the 204 participants, 89 (44%) were male and 115 (56%) were female. Their ages ranged from 18 to 32 years (M = 20.0, SD = 2.48). Each group contained an equal number (102) of randomly assigned participants. All responses were entered by hand into a computer database for analysis. All 204 questionnaires were filled out completely; that is, there were no missing data. According to the normal scoring procedure for the NEO-FFI, test items written in an introverted or non-conscientious direction (e.g., "I like to be alone.") were reverse scored so that high scores for each item indicated greater levels of extraversion or conscientiousness.

The scale reliabilities were high regardless of the construct and item order. All four of the reliability coefficients in this study (ranging from .76 to .82) were of the same magnitude as the coefficient alphas (.77 for extraversion and .81 for conscientiousness) reported by Costa and McCrae (1992) in the manual for the NEO-FFI.

The first hypothesis stated that a questionnaire with grouped items would have higher internal consistency reliability coefficients than a questionnaire with ungrouped items. Coefficient alpha for the extraversion items in the grouped format was .76; for extraversion items given in the ungrouped format, coefficient alpha was .80. The difference between these correlations was not significant, z = -.81, p > .05, and in the wrong direction; that is, the internal consistency of the extraversion scale was nonsignificantly higher for items presented in an ungrouped format. For the conscientiousness items, coefficient alpha was .82 for items in the grouped format and .80 for items in the ungrouped format. Although the results were in the hypothesized

8

direction, the difference between correlations was not significant, z = .51, p > .05. Given the null results for both significance tests, the first hypothesis was not supported.

The second hypothesis stated that a questionnaire with ungrouped items will have a higher scale intercorrelation (i.e., the correlation between the scores for each construct) than a questionnaire with grouped items. Total extraversion and conscientiousness scores were computed for each test taker. The correlation between scale scores for the items given in a grouped format was positive, r = .31. The correlation between scales for items given in an ungrouped format was also positive, r = .27. The difference between these correlations was not significant, z = ..31, p > .05, and in the wrong direction. Thus, the second hypothesis was not supported.

Discussion

The purpose of this study was to further examine the effect item grouping had on the internal consistency and scale intercorrelation of a personality test. I explored how item grouping affects test taker responses on a questionnaire measuring two dimensions of normal personality: extraversion and conscientiousness.

The results of this study did not offer any support for either of the hypotheses. The first hypothesis stated that a questionnaire with grouped items would have higher internal consistency reliability coefficients than a questionnaire with ungrouped items. Grouping items did not raise scale internal consistency. The second hypothesis stated that a questionnaire with ungrouped items would have a higher scale intercorrelation than a questionnaire with grouped items. Grouping items did not lower the scale intercorrelation.

It is difficult to infer why the results were not as hypothesized. One possibility is that there were not enough constructs in the questionnaire to cause a grouping effect. The inclusion of more constructs (such as agreeableness and openness) might offer more chances for test takers to become confused in the ungrouped condition which would reduce the quality of their responses. Another potential reason for the nonsignificant differences could be the number of items used per construct. This study used the 12-item per construct short form of the NEO-PI-R. As with the number of constructs, it is possible that including more items per construct would give test takers more chances to become confused in the ungrouped condition. Finally the ungrouped version of the questionnaire used an alternating item order (as opposed to a truly random item order). It is possible that an alternating order allows test takers to easily identify which items

10

measure which constructs and engenders a test taking process that closely matches a grouped order.

Conscientiousness and extraversion were chosen for this study because I believed they were different from each other, but not so different (e.g., math and verbal) that it would be overly simple for participants to differentiate between them. However, because the results did not yield any significant results, it is possible that conscientiousness and extraversion were too different to find an effect. There might have been an effect had the two constructs been more similar to each other, such as extraversion and agreeableness. As with the number of constructs, greater similarity should facilitate test taker confusion in the ungrouped condition.

Unfortunately, the results of this study mirror what previous studies have reported; that is, there is no clear pattern of results. As such, any conclusion that item grouping has an effect on test taker responses and the psychometric properties of a questionnaire would be, at best, premature.

References

- Allison, D. E. (1984). The effect of homogeneous vs. heterogeneous matching itemformat on test performance and reliability. *Alberta Journal of Educational Research*, 30, 91-94.
- Costa, P. T., & McCrae, R. R., (1992). Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual. Odessa, FL: Psychological Assessment Resources.
- Melnick, S. A. (1993). The effects of item grouping on the reliability and scale scores of an affective measure. *Educational and Psychological Measurement*, *53*, 211-216.
- Metzner, H., & Mann, F. (1953). Effects of grouping related questions in questionnaires. *Public Opinion Quarterly, 17,* 136-141.
- Schriesheim, C. A. (1981). The effect of grouping or randomizing items on leniency response bias. *Educational and Psychological Measurement, 41*, 401-411.
- Schriesheim, C. A., & DeNisi, A. S. (1980). Item presentation as an influence on questionnaire validity: A field experiment. *Educational and Psychological Measurement*, 40, 175-182.
- Schurr, K. T., & Henriksen, L. W. (1983). Effects of item sequencing and grouping in low-inference type questionnaires. *Journal of Educational Measurement*, 20, 379-391.

Appendix A

5

,

Informed Consent

INFORMED CONSENT DOCUMENT

Project Title: Item Grouping and the Effect on Factor Structure

Investigator: Julian T. Hunt, Department of Psychology, (270) 303-2051

You are being asked to participate in a project conducted through Western Kentucky University. The University requires that you give your signed agreement to participate in this project.

The investigator will explain to you in detail the purpose of the project, the procedures to be used, and the potential benefits and possible risks of participation. You may ask him any questions you have to help you understand the project. A basic explanation of the project is written below. Please read this explanation and discuss with the researcher any questions you may have.

If you then decide to participate in the project, please sign this form in the presence of the person who explained the project to you. You should be given a copy of this form to keep.

- 1. **Nature and Purpose of the Project:** The purpose of the current project is to examine differences in responses to a personality test based on how items are presented.
- 2. **Explanation of Procedures:** Upon signing this form, you will be given a short personality questionnaire that you should respond to in relation to yourself. Once you complete the questionnaire you will receive a more detailed explanation explaining the research and will be given an opportunity to ask questions about the research.
- 3. **Discomfort and Risks:** There are no discomfort or risks involved in this research.
- 4. **Benefits:** The results from this research will help to develop better testing methods.
- 5. **Confidentiality:** All questionnaires will be anonymous. You should not place any identifying information on the questionnaire itself.
- 6. **Refusal/Withdrawal:** Refusal to participate in this study will have no effect on any future services you may be entitled to from the University. Anyone who agrees to participate in this study is free to withdraw from the study at any time with no penalty.

You understand also that it is not possible to identify all potential risks in an experimental procedure, and you believe that reasonable safeguards have been taken to minimize both the known and potential but unknown risks.

Signature of Participant

Date

Witness

Date

Appendix B

Experimenter's Script

"Hello, my name is ______ and let me thank you for being a part of my study. I am a graduate student in the psychology program and I need your help in finishing my graduate thesis. The study will take only a few minutes of your time and the results will not affect you in any way. Furthermore, this is an anonymous study which means that once we start, you will not be asked for your name or any other identifying information."

Distribute the informed consent. "This form is called an informed consent. Please read over it while I explain a few issues. First off, your participation is voluntary. You should not be a part of this study if you do not want to. Additionally, you are free to withdraw at any point in time. That means that even if you agree to participate, you can stop, without penalty, at any time during the experiment. Finally, let me remind you that this study is anonymous. Except for this informed consent form, please do not put your name anywhere on the test materials. Are there any questions?"

Collect informed consent forms. Pass out the questionnaires, which should already be in alternating order in the stack.

"Here is the questionnaire – this is entire experiment, just 24 questions. Remember, don't put your name on here. But please answer honestly. For each question, please circle the response that best describes your opinion. When you are done, you may ______." (bring it up, wait for me – depending on the room)

Once the administration is complete. "Thank you for your time and attention. Are

there any questions that I can answer?"

Appendix C

Grouped Questionnaire

Age____ Sex: M or F

Please circle the response that best represents your opinion.

1.	I like to have a lot of peo Strongly Disagree	ple around me. <i>Disagree</i>	Neutral	Agree	Strongly Agree
2.	I laugh easily. Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
3.	I don't consider myself e Strongly Disagree	specially "light Disagree	t hearted". <i>Neutral</i>	Agree	Strongly Agree
4.	I really enjoy talking to p Strongly Disagree	eople. Disagree	Neutral	Agree	Strongly Agree
5.	I like to be where the act Strongly Disagree	ion is. <i>Disagree</i>	Neutral	Agree	Strongly Agree
6.	I usually prefer to do thin Strongly Disagree	ngs alone. Disagree	Neutral	Agree	Strongly Agree
7.	I often feel as if I'm burs Strongly Disagree	ting with energ Disagree	y. Neutral	Agree	Strongly Agree
8.	I am a cheerful, high-spin Strongly Disagree	rited person. Disagree	Neutral	Agree	Strongly Agree
9.	I am not a cheerful optim Strongly Disagree	iist. <i>Disagree</i>	Neutral	Agree	Strongly Agree
10.	My life is fast paced. Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
11.	I would rather go my ow Strongly Disagree	n way than be a Disagree	a leader of othe Neutral	rs. Agree	Strongly Agree
12.	I am a very active person Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
13.	I keep my belongings cle Strongly Disagree	an and neat. <i>Disagree</i>	Neutral	Agree	Strongly Agree
14.	I'm pretty good about pa Strongly Disagree	cing myself so <i>Disagree</i>	as to get things Neutral	done on time. <i>Agree</i>	Strongly Agree

15. I am not a very methodical person.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree		
16. I	try to perform all the ta	usks assigned to	me conscienti	ously.			
	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree		
17. I	17. I have a clear set of goals and work toward them in an orderly fashion.						
	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree		
18. I	waste a lot of time befo	ore settling dow	n to work.				
	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree		
19. I	19. I work hard to accomplish my goals.						
	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree		
20. When I make a commitment, I can always be counted on to follow through.							
	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree		
21. S	ometimes I'm not as de	ependable or re	liable as I shou	ld be.			
	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree		
22. I	am a productive persor	n who always g	ets the job done	e.			
	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree		
23. I never seem to be able to get organized.							
	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree		
24. I	strive for excellence in	everything I do).				
	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree		

Appendix D

Ungrouped Questionnaire

Sex: M or F Age____

Please circle the response that best represents your opinion.

1. I keep my belongings clean and neat.							
	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree		
2.	I like to have a lot of peor	le around me.					
	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree		
3	I'm pretty good about pac	ing myself so a	s to get things	done on time			
5.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree		
4	1 Llough ensily						
т.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree		
5	I am not a very methodica	1 nerson					
5.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree		
6	I don't consider myself es	necially "light]	hearted"				
0.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree		
7	7. I tru to perform all the tasks assigned to me conscientionally						
7.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree		
8 I really enjoy talking to neonle							
0.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree		
9. I have a clear set of goals and work toward them in an orderly fashion.							
2.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree		
10 Llike to be where the action is							
10.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree		
11. I waste a lot of time before settling down to work							
	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree		
12. I usually prefer to do things alone.							
1 22.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree		
13. I work hard to accomplish my goals							
10.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree		
14. Laften feel as if I'm bursting with energy							
14.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree		

15. When I make a commitment, I can always be counted on to follow through.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
16. I am a cheerful, high-spi	rited person.			
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
17. Sometimes I'm not as de	ependable or r	eliable as I sho	uld be.	
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
18. I am not a cheerful optim	nist.			
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
19. I am a productive persor	n who always g	gets the job do	ne.	
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
20. My life is fast paced.				
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
21. I never seem to be able t	o get organize	d.		
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
22. I would rather go my ow	n way than be	a leader of oth	ners.	
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
23. I strive for excellence in	everything I d	lo.		
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
24. I am a very active person	1.			
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree