DOCUMENT RESUME

ED 371 029 TM 021 660

AUTHOR Bode, Rita K.
TITLE Controlling for Demographic Characteristics in Person Measures Using a Many-Faceted Rasch Model.
PUB DATE Apr 94
NOTE 10p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Ability; Data Analysis; *Demography; *Estimation (Mathematics); Ethnic Groups; Grade 8; Group Membership; International Studies; Item Response Theory; Junior High Schools; Junior High School Students; Questionnaires; Racial Differences; Sex Differences
IDENTIFIERS Calibration; *FACETS Model; Person Parameters; *Rasch Model; Second International Mathematics Study

ABSTRACT

A benefit of using the multifaceted Rasch model is the ability to factor out or control for confounding factors in the estimation of person ability and item difficulty. This study experiments with a variation of the multifaceted Rasch analysis in calibrating the effects of demographic characteristics that are intended to overcome the problem of overparameterized person measures. The sample consisted of 1,319 U.S. eighth graders who participated in the Second International Mathematics Study (SIMS). The instrument was a seven-item measure of student effort taken from a questionnaire developed for the SIMS. In attempting to calibrate effects of the demographic characteristics, the FACETS program was not able to determine how much of person ability was due to the individual and how much was due to the gender or the racial/ethnic category. The proposed variation reverses the order in which the facets are calibrated, determines the gender and racial/ethnic effects and allocates the residual to person ability. This approach produces unambiguous person measures and, for these data, appears to make adjustments to the person measures for individuals based on their group membership. Two tables present study findings, and an appendix describes the variables. (Contains 5 references.) (SLD)

CONTROLLING FOR DEMOGRAPHIC CHARACTERISTICS IN PERSON MEASURES USING A MANY-FACETED RASCH MODEL

RITA K. BODE

UNIVERSITY OF ILLINOIS AT CHICAGO

Paper presented at the annual meeting of the American Educational Research Association, New Orleans, April, 1994

Controlling for Demographic Characteristics
in Person Measures using a Many-Faceted Rasch Model

There would appear to be interest in developing procedures for abstracting from measures of a variable those factors which confound the applicability of the measure to subsets of the population. Ordinarily, one would develop a measure for the instrument of interest on a heterogeneous sample and then conduct separate "bias" studies of the applicability of this measure on subsamples within the population. If it were possible to develop a measure that "controls" for differences in demographic characteristics that may affect the measure's applicability to members of these groups, the resulting measure would be a purer, more unambiguous estimate of the underlying variable.

Uses to date of FACETS, the multifaceted Rasch analysis (Linacre, 1989), have in general been focused on situations in which facets other than item difficulty and person ability such as rater severity and task difficulty are measured (Lunz, Wright, & Linacre, 1990; Engelhard, 1992). In such studies, ratings given to persons rated on a sample of tasks by a sample of raters are calibrated to determine the "net" person ability, regardless of which rater evaluated the performance and which tasks were sampled. The additional facets in these studies are characteristics of the task rather than person.

One of the major benefits of the use of a multifaceted Rasch analysis is the ability to "factor out" or "control for" confounding factors, such as rater severity and task difficulty, in the estimation of person ability and item difficulty. Estimates thus obtained for person ability, therefore, are independent of the specifics of the testing situation and need not be tied to a mistaken belief that these other factors don't affect the person estimates. Similarly, expanding this capability to variations across the person population would enable one to develop unambiguous person measures that represent the variable regardless of which demographic group one is a member. In situations in which differences in the performance of members of certain demographic groups can be attributed to bias in the measure, the ability to control for these differences would enable one to develop "bias-free" measures.

A recent use of FACETS has been to estimate the effects of demographic characteristics of persons by estimating item difficulty and person ability, anchoring the calibrations for these measures, and then adding demographic characteristics to the model (Mislevy, as cited in Linacre, 1993). In such a use, because the person ability and item difficulty measures are anchored, their calibrations are not allowed to change as a result of adding the demographic characteristics to the model. The additional facets in these studies, in contrast to the original studies of ratings, are characteristics of the person.

An extension of this methodology would be to measure person ability "controlling" for differences in the effects of the demographic characteristics. One such example is a study in which bias in college student ratings of instruction was detected and corrected (Haladyna & Hess, in press). In this study, faculty members were evaluated by students for whom certain demographic characteristics were available. The study found and corrected for significant variation in faculty ratings due to three non-evaluation facets: the gender of the rater, the course type, and whether or not the course was required for the rater.

Some attempts to perform such calibrations, however, have met with problems (Linacre, personal communication). When all facets are estimated, variation in person responses are over-parameterized and loosely connected subsets of elements are created in the measures of person ability. These loosely connected subsets occur because the amount of variation in the demographic characteristics is more than can be accommodated by the calibrated person measures. This outcome can occur whether or not person and item measures are anchored before estimating all facets.

This study experiments with a variation of the multifaceted Rasch analysis in calibrating the effects of demographic characteristics that are intended to overcome the problem of over-parameterized person measures. The development of such variations would expand the use of this procedure to situations in which it is now not possible because of excessive variation in the other person-related facets.

METHODOLOGY

While the applications of a multifaceted Rasch analysis have solved various measurement problems, a variation of these procedures may be needed in order to produce measures which would be unambiguous regardless of subset of a population to which a person was a member. It therefore, might be of interest to compare the measures obtained from the proposed variation to those obtained without taking demographic characteristics into account on a single dataset in order to determine what effect this variation has on the estimation.

Sample: The sample used in this study came from U.S. eighth-grade students who participated in the Second International Mathematics Study (SIMS). The sample consisted of 1319 students from typical (as opposed to remedial or enriched) eighth-grade mathematics classes for whom complete data were available for two administrations of various survey instruments.

Instrument: The instrument used in this study is a seven-item measure of student effort taken from a 54-item student questionnaire that was specifically developed for SIMS. The questionnaire was designed to measure students' attitudes toward mathematics; the instrument used in this study consisted of items which focused on wanting to do well in mathematics with some emphasis on the amount of effort needed to accomplish this goal. The instrument had previously been calibrated and was found to fit the Rasch model. Some of the items were negatively stated. The rating scale for these items was reversed and separate calibrations for positively and negatively stated items were included in the model. A copy of the survey items used in creating the effort measure is presented in Appendix A.

The demographic characteristics consisted of student gender and racial/ethnic category. A copy of the survey items measuring these demographic characteristics is also presented in Appendix A. The numbers in parentheses indicate the numbers of students in the sample in each of these demographic categories.

Analysis: The typical order in which facets are calibrated is to first calibrate the person and item measures and then add the other facets to the model. Previous analyses of these data in which all four facets were calibrated (with and without anchoring) produced unconnected subsets of elements for the person measure. In attempting to calibrate the effects of the demographic characteristics, FACETS was not able to determine how much of person ability was due to the person him/herself and how much was due to the gender and racial/ethnic category to which he/she belonged.

The variation in procedures proposed in this study reverses the order in which the facets are calibrated: first the item and step calibrations and the demographic facets are estimated and anchored and then the person facet is estimated. In this variation, the gender and racial/ethnic effects are determined and the residual is allocated to person ability. The resulting person ability was adjusted by the gender and racial/ethnic effects of whichever category to which the person belonged.

Three separate runs of FACETS were needed to conduct this analysis. The initial analysis was run to produce the best estimates of item calibrations based on only person and item data. Item and step calibrations were then anchored and the analysis of items, gender, and racial/ethnic data was run to produce the best estimates of the gender and racial/ethnic facets. Finally, item, step, gender, and racial/ethnic calibrations were anchored and the analysis of all facets was run to produce the best estimates of person effort.

The criteria used in evaluating the results of the variation as compared to initial analysis in which only item and person measures were estimated were a) whether the variation produced unambiguous person measures and b) the effect the variation had on person measures for students by gender and racial/ethnic category.

RESULTS

The summary of the final FACETS analysis using the variation of the calibration order is presented in Table 1. Person effort is essentially normally distributed. Males report expending slightly more effort and the racial ethnic effect is as follows: Latin American students report significantly higher levels of effort and Mexican American and Native American students report slightly lower levels than the remaining subgroups of students. It should be kept in mind that these are self-reported estimates of student effort which may vary from the actual effort expended in studying mathematics--a distinction which may be important in interpreting these results.

The map for the effort measure shows a good spread of item difficulty with one item, "I really want to do well in math," considerably easier to agree with than the rest of the items and those items dealing with spending time on mathematics are the hardest to agree with. The progression of this measure of effort then is: a) simply wanting to do well in mathematics, b) willingness to study math, c) trying hard in studying math, and finally, d) actually spending time in studying math.

The effect of taking gender and racial/ethnic category into account in calibrating effort measures for individual students is presented in Table 2. A student from each gender and racial/ethnic category who obtained the same effort measure (0.36) in the initial analysis in which the demographic characteristics are excluded was identified. The table consists of the effort measure obtained for these 14 students from the final analysis in which the demographic characteristics are included.

Table 2 shows the gender effect to be 0.15 logits; females average 0.15 logits lower than males so measures for females have been adjusted upward by that amount. The effect of controlling for racial/ethnic category is a downward adjustment in the effort measure for all subgroups. The measures for Latin American students (whose average effort measure is the highest) are adjusted the most and those for Mexican American and Native American students (whose average effort measures are the lowest) are adjusted the least, with the adjustments for the remaining subgroups somewhere in between.

Table 1

All Facets Summary Table

Measr	+Person	+Gender	+Racial/Ethnic		-Effort	
2	+ . (more effort)	+ (more effort)	+ (more effort)		+ (harder to agree with)	
	*.					
	.					
	.					
	*.					
	.					
	**.					
	.					
	**					
	***.					
1	+ .	+	+		+	
	**.					
	****.					
	.				will work long to understand new idea	
	****.	Male			spend a lot of time on math	
	*****.	Female	Latin American			
	****.					
	.					
	*********.				feel challenged by difficult problems	when I try hard I do well
	.		White			
0	* *********.	*	* Other		*	
	.		African American	Asian American	would study math if given choice	
	*********.		Mexican American	Native American	usually understand math class	
	.					
	********.					
	.					
	*****.					
	*****.					
	.					
	**.					
-1	+ ***.	+	+		+	
	.					
	*.					
	.				really want to do well in math	
	*.					
	*.					
	.					
	.					
	.					
-2	+ . (less effort)	+ (less effort)	+ (less effort)		+ (easier to agree with)	

Measr	* = 13	+Gender	+Racial/Ethnic	-Effort

Table 2

Adjusted Effort Measures for Persons with Same Initial Measure

	Female	Male
Native American	0.14	-0.02
African American	-0.01	-0.16
Mexican American	0.12	-0.03
Latin American	-0.62	-0.77
Asian American	-0.03	-0.18
White	-0.23	-0.38
Other	-0.08	-0.23

Discussion

Use of the variation in the order in which facets are anchored and estimated succeeded in producing unambiguous person measures. Thus, at least for these data, it appears that, in situations in which person ability is nested within the other person facets, the variation in the use of FACETS in which person measures are estimated after items and the demographic characteristics have been anchored produces the results that one would anticipate using FACETS. That is, FACETS appears to make adjustments to the person measures for individuals based on their group membership. This variation may not necessarily work in all situations; its applicability in other instances in which loosely-connected person measures have resulted would have to be determined.

The use of these adjusted measures treats the person effort measure as the residual after the effects of gender and racial/ethnic category are taken into account. It is similar to the use of demographic characteristics as covariates in either ANOVA or regression analyses. The difference is that the use of FACETS in this manner performs both the measurement and statistical analysis function in one step.

Regarding the use of person measures obtained using this variation, care should be taken as to the fairness of using adjusted versus nonadjusted measures; that is, whether the effect of demographic characteristics should be controlled. In this case, the use of adjusted effort measures has the effect of treating gender or racial/ethnic differences in reported effort as if they were biasing factors in the measures obtained that should controlled to provide an accurate measure of a student's reported effort. Use of rater severity and task difficulty are conceptually clear in terms of the desire to control. The situations in which the use of adjusted measures are appropriate and justification for controlling for demographic characteristics is a lot less clear.

In situations in which determining the effects of the demographic characteristics without the adjustment to take these characteristics into account are more appropriate, the two-step analysis used by Mislevy (as reported in Linacre, 1989) could be used. In the initial analysis, person and item estimates are calibrated; in the second, the calibrations for these two facets are anchored and all but the person measures are then estimated. In this second analysis, FACETS shows what the effect of each of the demographic characteristics is on the resulting person measure without making the adjustment. In that way one could detect the existence of significant effects of a "biasing" components of the measures but not correct for them.

References

Engelhard Jr., G. (1992). The measurement of writing ability with a many-faceted Rasch model. Applied Measurement in Education, 5(3), 171-191.

Haladyna, T., & Hess, R. (In press). The detection and correction of bias in student ratings of instruction. Review in Higher Education.

Linacre, J.M. (1989, 1993). Many-Facet Rasch Measurement. Chicago: MESA Press.

Linacre, J.M., & Wright, B.D. (1993). A user's guide to FACETS. Chicago: MESA Press.

Lunz, M.E., Wright, B.D., & Linacre, J.M. (1990). Measuring the impact of judge severity on examination scores. Applied Measurement in Education, 3, 331-345.

APPENDIX A

Description of Variables

EFFORT: Seven-item scale consisting of items from the posttest student questionnaire:

Tell, on a five-point scale, how much you agree with the feelings expressed in each statement below.

A	B	C	D	E
Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree

I feel challenged when I am given a difficult mathematics problem.
No matter how hard I try I still do not do well in mathematics. (When I try hard I do well in mathematics) *
I usually understand what we are talking about in mathematics class.
I will work a long time in order to understand a new idea in mathematics.
I really want to do well in mathematics.
I refuse to spend a lot of my own time doing mathematics. (I'm willing to spend a lot of my own time doing mathematics) *
If I had my choice I would not learn any more mathematics. (If I had my choice I would learn more mathematics) *

* negatively stated items reversed

GENDER: Dummy variable: 1 = Female; 2 = Male

Indicate your sex: boy (717)
girl (602)

RACIAL/ETHNIC CATEGORY: Categorical variable from 1-7

How do you describe yourself?

Native American or American Indian	(19)
Black or Afro-American	(90)
Mexican-American or Chicano	(45)
Puerto Rican or other Latin American	(21)
Oriental or Asian-American	(19)
White or Caucasian	(1065)
Other	(47)