

---

# METHODS AND TECHNIQUES

---

## Teaching Traits and Constructs through Scale Construction: A Need and a Strategy

Jeffrey D. Holmes  
*Ithaca College*

*Most students have some experience with purported measures of psychological constructs before entering college. Critical analysis of such measures requires familiarity with test development strategies. The activities outlined in this article allowed students to complete the steps involved in scale construction and empirical analysis in a relatively brief period of time. Steps included trait conceptualization, searching for literature, generating hypotheses, operationalizing constructs, collecting and analyzing data, and interpreting the quality of instruments. The activity took place in an introductory laboratory course but contains sufficient flexibility to make it useful in a variety of courses. Student evaluations of the activities were positive, and the article includes evidence of student learning.*

Students become familiar with a variety of reputed personality and cognitive ability tests through nonacademic sources such as the Internet and popular media. The wide dissemination of such scales may promote an unrealistically simplistic impression of construct definition and test construction. Without sufficient training in research methods and scale construction, students are not equipped to critically evaluate the quality and relevance of these measures. The professional-looking presentation of such scales, coupled with immediate feedback, may give the results an air of scientific credibility that is unwarranted and that fails to enhance students' understanding of the nature of testing; the difficulty of quantifying abstract psychological constructs; and the complex steps required to develop tests possessing adequate levels of reliability, validity, and utility. Several authors (e.g., Connor-

Greene & Greene, 2002; Peters & Sikorski, 1997) have noted that most Internet sources are not subject to any form of evaluation and are therefore more likely than peer-reviewed sources to contain material that is misleading, erroneous, or deceptive. Further, Beins (1993) cited evidence that students often do not engage in critical thinking with respect to research issues and concluded that students tend to accept personality test results without critical evaluation.

More than 20 years ago, Benjamin (1983) offered a method for teaching personality assessment concepts to undergraduates. He addressed strategies for involving students in construction of personality scales, collection of data, and basic statistical analyses. This approach is still applicable, and the widespread presence of computers and the increased accessibility of software for statistical analyses have reduced the practical complexities of data collection and analysis. This reality makes an update of this excellent strategy worthwhile. More recently, Dollinger (2004) provided the guidelines for a different but related activity focused specifically on the Big Five traits and designed to teach students about the prediction of specific behavioral tendencies from scores on more general personality scales. The articles by Benjamin and Dollinger provide excellent guidance for instructors wishing to engage students in active learning and critical thinking with respect to the use of personality scales. This article provides data to specifically illustrate the need for teaching scale construction to undergraduate psychology students and advocates for additional scale development and data analysis steps

to further enhance student sophistication with regard to measurement of traits and constructs. The activity, which followed from the data presented subsequently concerning student familiarity with measures from popular media sources, enhances students' understanding of psychological data by introducing them to the steps involved in construct measurement.

## Method

To assess student familiarity with popular measures of psychological constructs, 127 undergraduates enrolled in a freshman-level psychology lab course completed anonymous surveys concerning their experiences with such measures. Students reported their familiarity with tests from magazines and the Internet, the specific types and number of times they had completed such tests, and their impressions of the validity of the tests.

## Results

The vast majority of students (93%) reported having completed some type of psychological test in a magazine or on the Internet, with 48% having done so four or more times. Of those having completed scales from popular sources, 78% had completed personality measures, 31% had completed intelligence tests, and 8% could not remember the specific type of test. Perhaps most important, the data were consistent with past findings (Beins, 1993; Forer, 1949) on the Barnum effect in that the majority of students reported that the results they received seemed accurate (77%) and that the test seemed to measure what it was supposed to measure (80%). These data provide evidence that most students come to college having had some experience with scales ostensibly measuring various psychological constructs and that they place a great deal of confidence in the validity of these scales.

### The Scale Construction Activity

I developed this exercise for an introductory psychology laboratory course for the purpose of teaching a number of skills pertaining to construct measurement. The activities are relevant to a variety of courses including personality, research methods, and testing. The topics emphasized include personality theory and scale construction, as well as additional measurement issues such as reliability, validity, and the operationalization

of abstract constructs. The complete process includes development of a single-trait personality measure, collection of data, and examination of the associations between the new scale and existing instruments. The activity as presented requires 2 to 3 hr of class time and 2 to 3 hr of simple word processing and data entry, which the instructor might designate to a student assistant.

### *Session 1: Scale Construction*

The first step involved teaching about scale construction and facilitating development by students of a personality scale. During the first laboratory session, students discussed commonly researched personality traits and studied strategies for scale construction. The instructor assigned one of three traits (extraversion, openness to experience, or optimism) to each lab group. A laboratory assistant then guided the students as they searched Web-based literature from PsycINFO and other databases to better understand and conceptualize their trait.

Following this research and discussion, students each wrote at least two original items they believed would reflect the target trait. The students received specific instructions regarding the most effective way to write scale items (use of simple language and Likert scales, the perils of double-barreled questions, etc.). This process yielded at least 18 scale items (2 from each of 9 students). The members of the lab group then arrived at a consensus concerning the five items that they believed best represented the trait in question, and these items comprised the personality scale. I chose this number of items to keep the total number of items across all groups manageable and also to compel students to select what they believed to be the most precise items from those initially suggested. All laboratory sections repeated this process and developed unique scales. Multiple groups examined each trait, resulting in several versions of each trait scale. Use of this strategy permitted later analyses of both inter- and intratrait relationships. Each group spent approximately 90 min on the scale construction process.

### *Session 2: Data Collection*

After all laboratory groups had completed Session 1, a lab assistant combined the scales produced by each group into a single typed scale for purposes of administration. As part of a second laboratory meeting, all students responded to the items written by all lab sections, as well as several online single-trait personality measures including perfectionism, locus of control,

sensation seeking, and Type A behavior. These scales are useful because students can easily understand the constructs and make informed predictions of the associations with their scales. Depending on the goals of a particular course, an instructor might substitute other widely available trait scales for those listed. Because the completion of the scales was not time consuming, students spent the remaining laboratory time using PsycINFO to search for relevant literature to make predictions about how their trait scale would relate to other personality measures (instructors might assign this step as a homework activity to reduce demand for computers and class time). Each student retrieved two empirical articles reporting correlations between the assigned trait and other psychological constructs, and students used these articles to generate hypotheses and to focus their analysis and interpretation of the data.

### Session 3: Data Analysis and Interpretation

The final step in this exercise was the analysis of the personality data. A lab assistant compiled the data, and the students performed descriptive and correlational analyses appropriate for an introductory-level course. Students explored sex differences in the various traits, as well as the interrelations between similar and distinctive trait scales. One of the most educationally valuable results involved intratrait correlations. The correlations between different scales designed to measure the same trait were generally moderate, but there were some instances in which these measures were only weakly correlated with each other. For example, two measures of openness to experience correlated  $-.04$ , two measures of optimism correlated  $.17$ , and two measures of extraversion correlated  $.21$ , meaning that scales designed to measure identical traits had very little shared variance (0%, 3%, and 4%, respectively). These findings provided an excellent opportunity to discuss reliability and validity issues and the nuances of scale construction. The correlations also provided the groundwork for a more general discussion of construct operationalization by demonstrating that different researchers may conceptualize traits in remarkably different ways. Students used the intertrait correlations (between the student-written trait scales and measures of other traits) to test a priori hypotheses regarding the various traits. For example, students correctly predicted that extraversion would be positively associated with sensation seeking and openness to experience. The procedures culminated in each student writing a comprehensive laboratory report that included all elements of an empirical research manuscript. Students

received feedback on their reports based on a rubric addressing use of APA style, the reporting and interpretation of statistical analyses, and the use of existing literature for the development of research hypotheses.

## Evaluation and Adaptation to Other Courses

Informal student responses concerning this set of activities were quite positive. Hettich (1974) outlined the benefits of using student-generated data when teaching research methods, and this was one of the elements of the current activity that students found most interesting. Following the data analysis session in each of the first three semesters that I used this exercise, students completed a brief evaluation of the activities. Evaluations from 171 students were generally positive (see Table 1). Most students thought the activities fulfilled a variety of learning objectives, and more than half reported being more interested in personality research as a result of the exercises.

I developed this activity to increase students' awareness of and appreciation for the complexities of scale construction. Recognizing the limitations of using strictly attitudinal data to assess student learning, I used a procedure similar to that of Madsen, Van Abbema, Allen, and Schmidt (2006) and asked a second group

**Table 1. Students' Responses to Evaluation Survey**

Statement	Percentage Endorsing Either Agree or Strongly Agree <sup>a</sup>
The activities were effective for learning about personality research.	84
The activities were effective for learning about the development of personality scales.	87
The activities were effective for developing research skills.	76
The activities were effective for learning about correlational data analysis.	82
The activities made me more interested in the study of personality.	53
The activities should be used in future psychology lab courses.	84

<sup>a</sup> $n = 171$ .

**Table 2. Evidence of Student Learning**

Statement	Pretest <sup>a</sup>		Posttest		<i>t</i> (55)
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Writing a meaningful personality test requires many theoretical considerations.	4.02	.75	4.23	.66	2.27*
Writing a meaningful personality test requires many statistical considerations.	3.91	.64	4.16	.73	2.24*
Developing an operational definition is an important step when writing a personality test.	3.84	.80	4.46	.74	4.59*
Test reliability is an issue of concern when writing and evaluating personality tests.	4.02	.86	4.36	.92	2.42*
Test validity is an issue of concern when writing and evaluating personality tests.	4.18	.77	4.68	.51	4.39*

Note. Scale anchors of 1 (*strongly disagree*) and 5 (*strongly agree*).

<sup>a</sup>*n* = 56.

\**p* < .05 (posttest compared to pretest).

of students to respond to several items prior to beginning the activity and again immediately after completing the activity. Paired samples *t* tests demonstrated that on average, the students (*n* = 56) gained greater awareness of the importance of several critical scale development issues (see Table 2). In spite of range restriction due to the nature of the assessment items (i.e., students who did not know whether the issue was important would likely respond at the neutral midpoint of 3, leaving less room for change at posttest), student responses indicated significantly greater appreciation for a variety of important measurement issues and greater awareness of the complexity of the scale construction following the activity than at the outset. Students in this sample also indicated greater agreement with the statement, "Writing a personality test would be an interesting task" at the conclusion of the activity than they had at the beginning, *t*(55) = 2.13, *p* < .05, *r* = .28.

Finally, as a second and still more direct measure of student learning, I constructed a test consisting of nine difficult objective items addressing reliability, validity, operational definitions, and correlations. Members of the class during the most recent semester took the test prior to beginning the activity and again immediately after completing the activity. A paired samples *t* test demonstrated that the students (*n* = 48) scored sig-

nificantly higher at posttest (*M* = 6.35, *SD* = 1.54) than at pretest (*M* = 3.31, *SD* = 1.39), *t*(47) = 11.96, *p* < .001, *r* = .87.

Instructors can easily adapt this activity to make it appropriate for other courses. Although I designed it for introductory courses, it is flexible enough to work well in upper level courses through the inclusion of more sophisticated analyses such as multiple regression and through the exploration of moderating relationships. For example, instructors could introduce (a) regression by having students examine the incremental prediction of optimism using extraversion and openness to experience (or other included scales), (b) moderation by testing whether trait correlations (e.g., between extraversion and optimism) are moderated by sex, and (c) more complex reliability issues by including item-total correlations and internal consistency analyses to determine how various items fit with intended constructs. Given sufficient sample size, instructors also could introduce the topic of factor analysis by analyzing the factor structure of the students' items. Those teaching personality could choose from numerous alternative trait scales that are widely available. Instructors also could adapt the exercise to large lecture classes by combining the scale construction and data collection sessions. This would be possible by assigning background readings in advance and separating the students into small groups (analogous to the lab sections described earlier) to complete the scale construction steps during class. Before the students departed, the instructor could collect and quickly number the items before reading them aloud to the students, who would respond on their own paper. The students or instructor could then analyze the data outside of class and discuss the results during a subsequent lecture.

The method described adds to the work of Benjamin (1983) and Dollinger (2004) in several important ways. First, the method reported here allows students to explore both intra- and intertrait relationships by analyzing multiple measures of the same construct. Student discussions indicated that the often low correlations between different scales ostensibly measuring the same trait were surprising and thought-provoking for them. Second, the approach provides students with the experience of writing several novel items using a Likert format, thus permitting analyses ranging in complexity from basic descriptive statistics, correlations, significance levels, and regressions to more advanced procedures such as internal reliability and item analysis. Finally, the activity yielded a variety of outcome data demonstrating its effectiveness. Student evaluations of the activity were positive, and students demonstrated

greater knowledge of a variety of measurement issues following the activity than they had at the outset. The activity is therefore a useful and demonstrably effective strategy both for teaching the basic steps involved in scale construction and for raising student awareness of a variety of concerns relevant to psychological measurement.

## References

- Beins, B. C. (1993). Using the Barnum effect to teach about ethics and deception in research. *Teaching of Psychology, 20*, 33–35.
- Benjamin, L. T., Jr. (1983). A class exercise in personality and psychological assessment. *Teaching of Psychology, 10*, 94–95.
- Connor-Greene, P. A., & Greene, D. J. (2002). Science or snake oil? Teaching critical evaluation of “research” reports on the Internet. *Teaching of Psychology, 29*, 321–324.
- Dollinger, S. J. (2004). Predicting personality-behavior relations: A teaching activity. *Teaching of Psychology, 31*, 48–51.

- Forer, B. R. (1949). The fallacy of personal validation: A classroom demonstration of gullibility. *Journal of Abnormal and Social Psychology, 44*, 118–123.
- Hettich, P. H. (1974). The student as data generator. *Teaching of Psychology, 1*, 35–36.
- Madsen, S. D., Van Abbema, D. L., Allen, C. C., & Schmidt, R. E. (2006). Questioning claims of baby genius: Students evaluate advertisements of infant stimulation products. *Teaching of Psychology, 33*, 134–137.
- Peters, R., & Sikorski, R. (1997). Sharing information and interests on the Internet. *Journal of the American Medical Association, 277*, 1258–1260.

## Notes

1. An abbreviated version of this activity was presented as a poster at the Best Practices in the Teaching of Psychology Conference in Atlanta, Georgia, in October 2005.
2. I am grateful to Barney Beins, Hugh Stephenson, Sheila Singh, Randolph Smith, and the reviewers who provided helpful feedback on this article.
3. Send correspondence to Jeffrey D. Holmes, Department of Psychology, Ithaca College, Ithaca, NY 14850; e-mail: [jholmes@ithaca.edu](mailto:jholmes@ithaca.edu). Information on the rubric used to evaluate lab reports is available on request.

Copyright of Teaching of Psychology is the property of Lawrence Erlbaum Associates and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.