

Incremental Validity of the Frame-of-  
Reference Effect in Personality Scale Scores:  
A Replication and Extension

Mark N. Bing

Junior Authors: James C. Whanger,  
H. Kristl Davison, & Jayson B. VanHook

# Background

- Personality testing can be a major and successful component of a personnel selection system (Hogan, Hogan, & Roberts, 1996)
- Self-report questionnaires of the Big Five, especially of the Conscientiousness factor, have demonstrated criterion-related validity with measures of job performance and academic success (Schmidt & Hunter, 1998; Wolfe & Johnson, 1995)
- However, there are occasions when even Conscientiousness has demonstrated modest validity (Hurtz & Donovan, 2000; mean uncorrected validity coefficient = .11)

# Background

- As there is room for improvement, innovative changes to personality measures have been pursued to increase their predictive power (e.g., James, 1998; Schmit, Ryan, Stierwalt, and Powell, 1995)
- In their 1995 JAP article, Schmit et al. contextualized self-report personality items to situations that are relevant to the criterion of interest, and this contextualization provided respondents with a common frame-of-reference when answering test questions

# Background

- Why would one contextualize personality test items?
- Consider the noncontextualized item measuring the Achievement Striving facet of the Conscientiousness factor (NEO-PI-R) below:
  - I strive for excellence in everything I do.
- Given that a finite amount of cognitive resources can be devoted to responding, evaluating “everything” one does in life would probably lead to cognitive overload for most respondents
- Instead, respondents may focus on core self-evaluations when answering items, and these core self-evaluations may have different contextual bases

# Background

- Consider the possible alternative contextual bases for two different people:
  - Respondent Jack thinks of his devotion to his family and to his community when responding to the achievement striving item
  - Respondent Jill thinks of her devotion to school work and excellence in sports when responding to the item
- Thus, these respondents have different contextual bases and thus a different **Frame-of-Reference** when responding to this and other items

# Background

- Specifically, when responding to the item, Jack is thinking...
  - I strive for excellence in everything I do *in the community and with my family*.
- and Jill is thinking...
  - I strive for excellence in everything I do *at school and in sports*.
- Thus, our respondents could be answering different items due to their differences in item interpretations
- In sum, the different frames-of-reference of the various respondents would cause them to take different tests, and a high score for Jack would mean something different than a high score for Jill

# Background

- What are the implications?
  - 1. Differences in frame-of-reference lead to differences in item interpretation, and thus increases in measurement error and reductions in reliability
  - 2. Reductions in reliability lead to reductions in validity
  - 3. Thus, providing a common frame-of-reference that is relevant to the criterion of interest could standardize item interpretation and increase reliability and subsequent test validity
- This is exactly what Schmit et al. did with college students using grade point average (GPA) as a criterion

# Background

- In a between-subjects research design Schmit et al. administered traditional, noncontextualized Conscientiousness items and contextualized Conscientiousness items (NEO-PI-R) to college students
- One group responded to items like the following:
  - I strive for excellence in everything I do.
- Another group responded to items like:
  - I strive for excellence in everything I do at school.
- Schmit et al. also varied test instructions (applicant instructions vs. general instructions)

# Background

- Results: Scale scores composed of contextualized items had higher levels of reliability and validity than scale scores composed of noncontextualized items regardless of the instructional condition (i.e., applicant vs. general)
- However, because this innovative study utilized a between-subjects research design it left some questions unanswered
  - For instance, what is the correlation between contextualized and noncontextualized personality scale scores?
  - Can contextualized scales demonstrate incremental validity above and beyond noncontextualized scales, which is to say, “Are contextualized scales value-added for selection settings?”

# Background

- Additionally, because college GPA is highly correlated with cognitive ability, there is a classic unmeasured variables problem, and it is legitimate to ask, “Can the contextualized scale obtain incremental validity above and beyond BOTH the noncontextualized scale and cognitive ability?”
- The present study was designed to answer these questions via replicating and extending the Schmit et al. study

# Method

- Subjects
  - 308 entering freshman students at a small Southeastern university
- Research Design
  - Within-subjects factor: School contextualized vs. noncontextualized items from the NEO-PI-R Conscientiousness scale (order counterbalanced)
  - Between-subjects factor: Applicant vs. general instructions
  - Covariate: Cognitive ability as measured by ACT scores
  - Criterion: Cumulative college GPA
- Procedure
  - Conscientiousness scales administered first week of college during orientation; GPA collected at end of freshman year

# Results

- Applicant instructional condition:
  - $\alpha = .94$  for school contextual Conscientiousness
  - $\alpha = .93$  for noncontextual Conscientiousness
    - Although the difference in reliability is slight, it is significant,  $t(140) = 2.00, p < .05$
  - $r = .89$  ( $p < .01$ ) between contextual and noncontextual Conscientiousness under applicant instructions
  - $r = .46$  ( $p < .01$ ) between contextual Conscientiousness and college GPA
  - $r = .39$  ( $p < .01$ ) between noncontextual Conscientiousness and college GPA
    - This difference in validity is significant,  $t(141) = 2.00, p < .05$

# Results

- General instructional condition:
  - $\alpha = .95$  for school contextual Conscientiousness
  - $\alpha = .92$  for noncontextual Conscientiousness
    - Although the difference in reliability is again slight, it is significant,  $t(159) = 6.86, p < .01$
  - $r = .90$  ( $p < .01$ ) between contextual and noncontextual Conscientiousness under general instructions
  - $r = .51$  ( $p < .01$ ) between contextual Conscientiousness and college GPA
  - $r = .42$  ( $p < .01$ ) between noncontextual Conscientiousness and college GPA
    - This difference in validity is significant,  $t(161) = 2.98, p < .01$

# Results

- In spite of this high collinearity between contextualized and noncontextualized scale scores, incremental validity was obtained for the contextualized scale above and beyond BOTH the noncontextualized scale and ACT scores in each instructional condition (applicant and general)
- Applicant,  $\Delta R^2 = .05$ ,  $p < .01$ , Total  $\Delta R^2 = .44$ ,  $p < .01$
- General,  $\Delta R^2 = .07$ ,  $p < .01$ , Total  $\Delta R^2 = .46$ ,  $p < .01$

# Conclusions

- Increases in reliability resulted from providing respondents with a common frame-of-reference, and this was especially evident at the subscale level (i.e., with Competence, Dutifulness, etc.), as coefficient alpha is biased by number of items
- Increases in validity resulted from providing a common frame-of-reference
- School contextualized Conscientiousness obtained incremental validity in the prediction of GPA above and beyond BOTH noncontextualized Conscientiousness and cognitive ability

# Implications

- Both the current study and Schmit et al.'s study indicate that providing a common frame-of-reference to personality items enables individuals to make more specific responses, thus allowing them to present themselves more accurately
- Providing a common frame-of-reference increases reliability and validity via reducing measurement error
- This increase in personality test validity should lead to more accurate hiring decisions
- Future studies should test this possibility directly among working adults in organizational settings

Running Head: INCREMENTAL VALIDITY OF THE FRAME-OF-REFERENCE EFFECT

Incremental Validity of the Frame-of-Reference Effect in Personality Scale Scores:  
A Replication and Extension

Mark N. Bing

Naval Submarine Medical Research Laboratory

Junior Authors: James C. Whanger, H. Kristl Davison, & Jayson B. VanHook

***Winner – 26<sup>th</sup> International Student Paper Competition Sponsored by the International Personnel Management Association Assessment Council (IPMAAC), Invited Address on Content of Paper at IPMAAC's 26<sup>th</sup> Annual Conference, New Orleans, LA – Summer 2002***

***Under Revise-and-Resubmit Status at the Journal of Applied Psychology***

## Abstract

Context-specific personality items provide respondents with a common frame-of-reference unlike more traditional, noncontextual personality items. The common frame-of-reference provided by context-specific items helps to standardize item interpretation, and has been shown to reduce measurement error while increasing validity in comparison to noncontextual items (Schmit, Ryan, Stierwalt, & Powell, 1995). Although the frame-of-reference effect on personality scales scores has been well investigated (e.g., Schmit, et al., 1995), the ability of this innovation to obtain *incremental* validity above and beyond the well-established, noncontextual personality scale scores has yet to be determined. As there are substantive reasons that suggest high redundancy (i.e., multicollinearity) should be found between context-specific personality items and their noncontextual counterparts, the demonstration of *incremental* validity for this innovation should be undertaken prior to drawing the conclusion that this innovation is truly value-added for personnel selection. The current study replicates and extends work by Schmit, et al. (1995) to determine the value-added nature of the frame-of-reference effect in personality scale scores. The results indicated that context-specific personality items do indeed obtain incremental validity above and beyond both noncontextual items and cognitive ability. The implications of these findings for personnel selection are discussed.

## Incremental Validity of the Frame-of-Reference Effect in Personality Scale Scores:

## A Replication and Extension

A substantial amount of evidence suggests that cognitive ability is highly predictive of both job performance and academic success (Camara & Echternacht, 2000; Dollinger & Orf, 1991; Hunter & Hunter, 1984; Pearlman, Schmidt, & Hunter, 1980; Schmidt, Gast-Rosenberg, & Hunter, 1980; Schmidt, Hunter, & Caplan, 1981; Sibert & Ayers, 1989). There is also evidence that the Big Five personality factor of Conscientiousness is often an effective predictor of job performance and academic success (Barrick & Mount, 1991; Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; Schmit, Ryan, Stierwalt, and Powell, 1995; Tett, Jackson, & Rothstein, 1991). Conscientiousness has been characterized as encompassing personal qualities such as competency, dependability, self-discipline, and deliberation (Costa & McCrae, 1989). While Conscientiousness has been generally shown to contribute significantly to the prediction of performance over and above cognitive ability (e.g., Schmidt & Hunter, 1998), the prediction of job performance by this Big Five factor is often modest (see Hurtz & Donovan, 2000).

One reason for this limitation may be due to the measurement of personality traits in a rather general, noncontextual manner. For instance, most personality measures attempt to assess global personality constructs that are cross-situationally consistent rather than situationally-specific (Schmit et al., 1995). The theory of conditional dispositions (Wright & Mischel, 1987) suggests that the manifestation of personality traits (i.e., dispositions) is conditional upon certain situations, such that individuals may respond in one way in one situation and in a very different manner under a different situation. For example, an individual may behave in an extroverted manner at work, yet fail to display extroverted behaviors in non-work situations. Noncontextualized items that do not take the conditional nature of personality into account may

create problems for personnel selection insofar as these items ask individuals about their behaviors *in general*, and thus may be suboptimal in predicting behaviors *in specific contexts*, such as the workplace.

Noncontextual items are open to interpretation by respondents in comparison to context-specific items. When answering test items, one respondent may consider the way he behaves at work, and another may consider the way he behaves in social situations, and thus these respondents are in essence not responding to the same item *when taking into account their differences in item interpretation*. Such differences in item interpretation across respondents would lead to increases in measurement error and a subsequent reduction in item validity. Thus, one possibility for improving the validity of personality tests may lie in providing individuals with a common frame-of-reference for personality test items.

Schmit, Ryan, Stierwalt, and Powell (1995) examined this possibility by rewriting noncontextual self-report personality items into school-specific contexts. This *frame-of-reference effect* on item validity was investigated in Study 2 of their research by examining the relationship between the NEO-PI-R Conscientiousness scale (Costa & McCrae, 1992; Costa, McCrae, & Dye, 1991) and college Grade Point Average (GPA). In their well-conducted study, Schmit et al. (1995) used a between-groups research design with two factors to create four study conditions. The first factor, instruction type (*general* vs. *applicant*), was manipulated such that participants received either the original scale instructions or instructions to complete the scale as if applying for admission to a desired university. The second factor, item type (*noncontextual* vs. *school-specific*), was manipulated by administering either the original, unaltered NEO-PI-R Conscientiousness scale, or the Conscientiousness scale with items rewritten into a school context. For example, for the school-contextualized Conscientiousness scale the item “I strive for

excellence in everything I do” was contextualized by changing it to “I strive for excellence in everything I do at school.”

From a theoretical perspective, Schmit et al. (1995) pitted a combination of the self-presentation theory of item responding (e.g., Hogan, 1991; Leary & Kowalski, 1990) and the theory of conditional dispositions (Wright & Mischel, 1987) against socially desirable responding theory (e.g., Hough et al., 1990; Ones, Viswesvaran, & Reiss, 1996). Specifically, self-presentation theory suggests that when individuals respond to personality test items they are guided by abstract self-concepts and that their responses will be consistent with both how they see themselves and how they wish others to see them (see e.g., Taylor, Carithers, & Coyne, 1976). The integration of self-presentation theory with the theory of conditional dispositions suggests that providing a specific context for personality test items will increase respondents' ability to present themselves in a manner that is consistent with their behavior in specific situations. For example, individuals may behave differently in school vs. nonacademic situations, and providing a school context for personality items would permit individuals to respond in a manner that is more accurate in terms of their behaviors at school. Thus, self-presentation theory and the theory of conditional dispositions together argue that providing a common frame-of-reference for personality items would increase item validity by allowing greater accuracy in item responses and thus better measurement of the relevant personality trait.

In contrast, socially desirable responding theory suggests that greater specificity in personality items permits respondents to deduce the more appropriate or “socially desirable” response. Thus, such items might be easier to fake. In addition, providing specificity in instructions might provide respondents with motivation to answer in a socially desirable manner. For example, providing applicant instructions might provide an incentive to fake (e.g., to obtain

school admission), and the increased item specificity might make faking or response distortion easier. Thus, according to socially desirable responding theory, providing a common frame-of-reference would reduce validity by introducing item responding error and perhaps range restriction in scale scores.

Schmit et al.'s (1995) results supported the combined theories of conditional dispositions and self-presentation. They found that context-specific (i.e., school-specific) personality items increased the criterion-related validity of Conscientiousness scale scores in predicting college GPA. Instruction type, however, failed to produce a consistent effect on validity.

One limitation of Schmit et al.'s (1995) study is essentially an unmeasured variables problem—their criterion, college GPA, is highly correlated with cognitive ability, yet cognitive ability was not controlled for in their study. It is possible that the greater validity of context-specific items may be attributable to individual differences in cognitive ability, rather than to improved measurement of school-related Conscientiousness. Specifically, if by contextualizing items the “correct” or desirable answer becomes obvious to those higher in cognitive ability, but remains unclear to those lower in cognitive ability, then school contextualized items may incidentally measure cognitive ability whereas noncontextualized items would not have this redundancy with cognitive ability. Thus, if the variance in GPA attributed to item context-specificity overlaps substantially with the variance attributable to cognitive ability, then the increased validity attributed to a common frame-of-reference effect may instead be due to cognitive ability. In other words, the increased validity *would not* be directly attributable to a more precise measure of Conscientiousness, but rather to the intellectual ability to identify the cues indicating the *correct or desirable* answer.

Stated alternatively, the ability to *effectively distort* self-report personality (i.e., respond in a socially desirable manner) also may be related to cognitive ability. Specifically, the context may provide cues that enable individuals to infer what the criteria for success are, and that then will allow these individuals to detect the answers most congruent with success. In this case, providing a common frame-of-reference creates a more transparent *ideal* answer for the personality test-taker, but the degree of transparency will be a function of the cognitive ability of the test-taker, such that individuals higher in cognitive ability are more capable of inferring criteria for success and thus are more likely to deduce the correct answer. If this is true, then the increased validity found by Schmit et al. (1995) is more likely a function of socially desirable responding that is based on higher levels of cognitive ability rather than the result of more accurate self-presentation engendered by providing the test-takers with a common frame-of-reference. If context-specific items fail to add incremental validity above and beyond cognitive ability, then it is not possible to rule out cognitive ability as an explanation for the higher validity of context-specific personality items observed by Schmit et al. (1995).

In addition to the possible unmeasured variables problem, there is another reason for replicating the Schmit et al. (1995) study. The between-groups nature of the research design in Schmit et al.'s second study did not allow them to examine whether or not context-specific personality items could add *incremental validity* to the prediction of performance above and beyond noncontextual personality items. A number of researchers have suggested that a new measure should demonstrate incremental validity above and beyond established measures to be deemed pragmatic and value-added in application (see Cronbach & Gleser, 1957; Day & Silverman, 1989; Sechrest, 1963). Both cognitive ability (Dollinger & Orf, 1991; Hunter & Hunter, 1984; Pearlman, Schmidt, & Hunter, 1980; Schmidt, Gast-Rosenberg, & Hunter, 1980;

Schmidt, Hunter, & Caplan, 1981) and the Big Five factor of Conscientiousness (Barrick & Mount, 1991; Hough, et al., 1990; Tett, Jackson, & Rothstein, 1991) have been shown to be effective predictors of performance. Thus, the primary research question in the current study is whether *context-specific* personality items add to the prediction of performance above and beyond both cognitive ability and *noncontextual* personality items. Contextualized personality traits should, theoretically, be related to their noncontextualized counterparts. For example, school-specific Conscientiousness may very well extend from a more context-independent form of Conscientiousness, and as such school-specific Conscientiousness should correlate with noncontextual Conscientiousness. Given this plausible redundancy with the noncontextual personality items already in use in selection settings, it is important for context-specific personality items to demonstrate incremental validity above and beyond the well-established, noncontextual items if they are to be deemed useful for personnel selection.

Our hypotheses regarding the effects of context-specific items and instructions are consistent with the work of Schmit et al. (1995), except with regard to a slight difference in the rank-ordering of validities. Schmit et al. (1995) proposed that the rank-order of validities from highest to lowest would be: (1) applicant—school-specific, (2) general—school-specific, (3) general—noncontextual, and (4) applicant—noncontextual. We propose that self-presentation theory when considered in conjunction with the theory of conditional dispositions would reverse the predicted rank-order of the latter two validities, such that applicant instructions would lead to higher validity for noncontextual items in comparison to general instructions. Specifically, the applicant instructions condition creates a more consistent frame-of-reference for respondents than the general instructions condition, regardless of whether items are school-specific or noncontextualized. Applicant instructions should allow respondents to present themselves as

they would like to be perceived, more so than general instructions. Therefore, the validity under applicant instructions would be expected to be higher than under the general instructions, for each item type. Further, validities for scales composed of school-specific items, because they provide a common frame-of-reference for respondents and thus reduce differences in item interpretations across respondents, should have higher validities than scales composed of noncontextual items under both instruction types. The current study examined these predicted relationships, with a focus on the incremental validity of context-specific personality items above and beyond cognitive ability and noncontextual personality items.

## Method

### *Participants*

The sample consisted of 342 entering freshman students (approximately 62% female) at a small Southeastern university. The students participated on a voluntary basis in exchange for extra credit. Of these participants, 62% were between the ages of 17 and 18, 28% were between 19 and 20, 2% were between 21 and 22, 1% were 23 and older, and 7% did not indicate their age. The sample included 185 participants in the general instructions condition and 157 participants in the applicant instructions condition. Cumulative GPA was subsequently obtained for 164 participants in the general instructions condition and 144 participants in the applicant instructions condition, resulting in 308 cases for criterion-related analyses.

### *Measures*

*NEO-PI-R Conscientiousness Scale.* During the first week of the fall semester all study participants completed both the original Conscientiousness scale from the NEO-PI-R (Costa & McCrae, 1992), and the school-contextualized version used in the Schmit et al. study (1995). Order of scales was counter-balanced within instruction conditions, such that half of the

participants of each instruction condition received the noncontextual items first followed by school-specific items (*Order 1*), and the other half received the school-specific items first followed by the noncontextual items (*Order 2*).

Participants responded to items on a 5-point Likert scale ranging from 0 (*strongly disagree*) to 4 (*strongly agree*), and scale scores were computed such that higher scores indicated higher levels of Conscientiousness. In the current study, coefficient alpha was .93 for the noncontextual items and .95 for the school-specific items.

*General vs. applicant instructions.* Participants were randomly assigned to receive either general or applicant instructions for completing the Conscientiousness measures. The general instructions asked participants to answer the questions as directed on the original version of the Conscientiousness scale, whereas the applicant instructions were those used by Schmit et al. (1995) and asked participants to answer as if their scores would be used to determine admission to a desired university.

*Cognitive Ability.* The participants' scores on the ACT were used as a measure of cognitive ability. Numerous studies have demonstrated the validity of using college admission exams such as the ACT and SAT in predicting college performance (e.g., Camara & Echternacht, 2000; Sibert & Ayers, 1989). ACT and SAT scores were obtained from the university registrar, with the participants' permission and informed consent. When only the SAT score was available for a participant that score was converted to an equivalent ACT score via conversion tables provided by Educational Testing Services.

*Criterion.* The criterion measure was cumulative college grade point average (GPA) based on a 4.0 scale, with higher GPAs indicating higher course performance. Participants signed an informed consent form that gave the researchers permission to access their cumulative GPA

from the university registrar. GPA included grades for all courses completed through the end of participants' freshman year.

### *Procedure*

Participants were administered both contextual and noncontextual Conscientiousness scales with either the general or applicant instructions at the beginning of the semester in freshman orientation seminars. Thus, this study incorporated both between-subject and within-subject factors. The between-subject factor was type of instructions (general vs. applicant). The within-subjects factor was the counterbalanced presentation of noncontextual and school-specific Conscientiousness scales.

As in the Schmit et al. (1995) study, a \$10 prize was offered to the top scorers in the applicant instructions condition in order to motivate participants. With participants' permission, freshman year cumulative GPA and ACT (and SAT) scores were obtained from the university registrar at the end of the spring semester.

### Results

Table 1 presents the means, standard deviations, and coefficient alpha estimates of reliability for the study variables under both general and applicant instruction conditions. As the table shows, the reliabilities were similar across conditions and within acceptable levels. Means and standard deviations were comparable across conditions, although the mean for the school-specific Conscientiousness scale in the applicant instructions condition was significantly higher than the mean for the school-specific Conscientiousness scale in the general instructions condition ( $t(340) = 2.08, p < .05$ ).

Correlations among the study variables for each of the instruction conditions are shown in Table 2. In both instruction conditions, the noncontextual and school-specific

Conscientiousness scales are highly correlated. In the general instructions condition the correlation between school-specific and noncontextual Conscientiousness was  $.90$  ( $p < .01$ ), and in the applicant instructions condition the correlation was  $.89$  ( $p < .01$ ). Thus, as expected, there was a high degree of multicollinearity between school-specific and noncontextual Conscientiousness. Although correlations between ACT and both Conscientiousness scales were elevated under the applicant instructions condition in comparison to the general instructions condition, which would be predicted by socially desirable responding theory, these differences were not significant. Additionally, contextualizing the items to school settings clearly did not result in increased correlations between school-specific Conscientiousness and ACT in comparison to the correlations between noncontextual Conscientiousness and ACT under either instruction condition.

The validities for Conscientiousness with college GPA across each of the four conditions differ slightly from those obtained by Schmit et al. (1995). The highest criterion validity coefficient was obtained in (1) the general—school-specific condition ( $r = .51, p < .01$ ), followed by (2) the applicant —school-specific condition ( $r = .46, p < .01$ ), then (3) the general—noncontextual condition ( $r = .42, p < .01$ ), and finally (4) the applicant—noncontextual condition ( $r = .39, p < .01$ ). The significant validity coefficient obtained for the noncontextualized Conscientiousness scale in the applicant instructions condition (i.e.,  $r = .39$ ) contrasts with the findings of Schmit et al. (1995), who did not obtain a significant correlation between the noncontextual Conscientiousness scale and GPA in the applicant instructions condition.

In contrast to Schmit et al.'s (1995) results, the current study found that general instructions with school-specific items produced the highest validity, followed by applicant instructions with school-specific items. The pattern was the same for the noncontextual items.

Further, a test of significance of the difference between dependent correlations (Cohen & Cohen, 1983, pp. 56-57) revealed that the validity coefficient for the school-specific Conscientiousness scale ( $r = .46$ ) was significantly higher than the validity coefficient ( $r = .39$ ) for the noncontextual Conscientiousness scale within the applicant instructions condition ( $t(141) = 2.00$ ,  $p < .05$ ). The same test revealed a significant difference between validities for the school-specific Conscientiousness scale ( $r = .51$ ) and the noncontextual Conscientiousness scale ( $r = .42$ ) within the general instructions condition ( $t(161) = 2.98$ ,  $p < .01$ ). When considered together, the above results provided strong support for the operation of self-presentation theory (Hogan, 1991) at the item level insofar as item contextualization increased, rather than decreased, validity. The results for instruction level findings did not provide clear support for either self-presentation theory or socially desirable responding theory.

According to our interpretation of self-presentation theory when considered with the theory of conditional dispositions, the applicant instructions condition was expected to create a more consistent frame-of-reference than the general instructions condition, which would lead to higher validities for applicant instructions in comparison to general instructions. However, general instructions produced higher validities than applicant instructions for both school-specific ( $r = .51$  vs.  $r = .46$ ) and noncontextual ( $r = .42$  vs.  $r = .39$ ) Conscientiousness scales, although these differences were not significant ( $z = .57$ ,  $p = .29$ , and  $z = .31$ ,  $p = .38$ , respectively). Thus, neither self-presentation theory nor socially desirable responding theory was supported at the instruction level.

Hierarchical linear regression analyses were used to test the ability of context-specific personality items to demonstrate *incremental validity* above and beyond cognitive ability and traditional, noncontextual personality items. Analyses were conducted separately for the general

and applicant instructional conditions, and then for the full sample with instructional condition as a dummy-coded covariate. Order of entry of predictors into the hierarchical regression analysis was determined a priori, as suggested by Cohen and Cohen (1983). ACT was entered in the first step because cognitive ability is an extremely well-established predictor of both academic and work performance (e.g., Hunter & Hunter, 1984; Schmidt & Hunter, 1998). Order of item-type (i.e., contextual prior to noncontextual and vice-versa, dummy-coded) was entered in the second step, because it could potentially affect responding to Conscientiousness scale items.

Noncontextual Conscientiousness was entered in the third step because it is also a well-established predictor of both academic and work performance (e.g., Barrick & Mount, 1991; Tett, Jackson, & Rothstein, 1991; Wolfe & Johnson, 1995). Finally, school-specific Conscientiousness was entered in the fourth step, to examine the incremental validity of item contextualization in the prediction of GPA. When the full sample was used, instruction type was entered second, prior to order of item-type.

The results of the incremental validity analysis are shown in Tables 3 and 4. As expected, ACT scores were a significant, major predictor of cumulative GPA in both the general ( $\Delta R^2 = .26, p < .01$ ) and applicant instructions conditions ( $\Delta R^2 = .32, p < .01$ ), as well as in the full sample ( $\Delta R^2 = .29, p < .01$ ). Instruction condition did not contribute significantly to prediction in the full sample ( $\Delta R^2 = .00, p = .64$ ). Order of item type also failed to contribute significantly to the prediction of GPA in either condition (general,  $\Delta R^2 = .01, p = .25$ ; applicant,  $\Delta R^2 = .00, p = .47$ ) or the full sample ( $\Delta R^2 = .00, p = .17$ ). However, noncontextual Conscientiousness added substantially to the prediction of GPA beyond ACT in all analyses (general,  $\Delta R^2 = .12, p < .01$ ; applicant,  $\Delta R^2 = .06, p < .01$ ; full sample,  $\Delta R^2 = .09, p < .01$ ). Finally, regarding the primary question of interest, school-specific Conscientiousness added a significant amount of incremental

validity beyond that due to cognitive ability and noncontextual Conscientiousness in both general ( $\Delta R^2 = .07, p < .01$ ) and applicant ( $\Delta R^2 = .05, p < .01$ ) instructional conditions and in the full sample ( $\Delta R^2 = .06, p < .01$ ). Clearly, these results support the view that increases in item-context specificity represent a potentially value-added innovation of the frame-of-reference effect for personnel selection purposes.

It is relevant to note that the relationship between noncontextual Conscientiousness and GPA changes from positive to negative when school-specific Conscientiousness is added to the regression equation. Prior to entry of school-specific Conscientiousness, noncontextual Conscientiousness is positively related to GPA in both instructional conditions (general,  $\beta = .35, p < .01$ ; applicant,  $\beta = .26, p < .01$ ) and in the full sample ( $\beta = .31, p < .01$ ), such that higher Conscientiousness is related to higher GPA. However, when school-specific Conscientiousness enters the regression equation, the standardized regression weight for noncontextual Conscientiousness becomes negative and drops to nonsignificance for the general and applicant instruction conditions ( $\beta = -.25, p = .09$ , and  $\beta = -.23, p = .14$ , respectively), although it remains significant in the full sample ( $\beta = -.22, p < .05$ ). This is likely due to the high level of multicollinearity between the two Conscientiousness scales, and reflects the dominance of school-specific Conscientiousness in the prediction of GPA. Additionally, once school-specific Conscientiousness has been partialled from noncontextual Conscientiousness, increases in noncontextual Conscientiousness reflect increases in conscientious behaviors (e.g., orderliness, achievement-striving) devoted to nonacademic pursuits, and could thus detract from academic performance and result in the observed negative relationship with GPA.

## Discussion

These results support the findings of Schmit et al. (1995) that indicate increases in personality item-context specificity improve the criterion-related validity of personality scale scores. The contribution of the current study lies in the demonstration of the incremental validity of contextualized personality scale scores beyond both cognitive ability and traditional, noncontextual personality scale scores. This incremental prediction was obtained in spite of entering the context-specific Conscientiousness scale into the regression equation in the final step, thus eliminating the possibility that the increased validity observed for context-specific scale scores could be explained by relations between context-specific Conscientiousness and either cognitive ability or noncontextual Conscientiousness.

Although the rank order of validity coefficients was similar to that obtained by Schmit et al. (1995), there are a few noteworthy differences in our results. First, the validity for the general instructions—school-specific condition was higher than that of the applicant instructions—school-specific condition, whereas Schmit et al. obtained the reverse order. However, this difference was nonsignificant in the current study as well as in the Schmit et al. study. The difference between general and applicant instructions for the noncontextual items also was nonsignificant, as Schmit et al. found. Nonetheless, we obtained significant correlations between Conscientiousness and GPA in the applicant instructions—noncontextual condition, in contrast to the nonsignificant finding obtained by Schmit et al. (1995). This is not surprising, given that Conscientiousness is one of the strongest predictors of performance (Barrick & Mount, 1991).

Additionally, we found significant differences between school-specific and noncontextual validities in both the general instructions condition, as well as in the applicant instructions condition, whereas Schmit et al. found a significant difference only under applicant instructions.

Our results thus provide strong support for the proposition that providing contextualized items increases validity, as predicted by the self-presentation theory of item responding. This has important implications for personnel selection, as it suggests that when screening job candidates for a position, contextualized items (e.g., work-specific items) are potentially more useful in the prediction of job performance than noncontextual items. Therefore, contextualized personality scale scores should significantly improve personnel selection decisions.

The failure of applicant instructions to produce higher validities than general instructions was in contrast to our interpretation of self-presentation theory when considered with the theory of conditional dispositions. It was expected that providing a specific context in the instructions would also provide a common frame-of-reference across respondents and thus increase validity. However, general instructions showed higher validity than applicant instructions, although these differences were not significant.

A noteworthy finding was the reversal in sign of the standardized regression weight for noncontextual Conscientiousness in the hierarchical regression analyses. Prior to the entry of school-specific Conscientiousness, noncontextual Conscientiousness was positively related to GPA, i.e., increases in noncontextual Conscientiousness predicted increases in GPA. However, in the presence of school-specific Conscientiousness, noncontextual Conscientiousness had a negative relationship with GPA. Given that an individual has a finite amount of time and energy to devote to various activities, an individual who is conscientious at school will presumably choose to devote a substantial amount of time to school-relevant activities. However, as the individual becomes more conscientious in general (i.e., higher in noncontextual Conscientiousness), he or she may devote greater amounts of time to non-school activities and less time to school-related activities, leading to decreases in GPA. In contrast, an individual who

is conscientious at school but not as conscientious in general (i.e., lower in noncontextual Conscientiousness) may devote less time to outside activities and thus have greater time to invest in school-related activities that result in achieving a higher GPA. Therefore, when school-specific Conscientiousness is controlled for, higher noncontextual Conscientiousness may lead to lower GPA via drawing the devotion and achievement-striving of the individual into non-school activities that take time away from school-relevant, GPA-enhancing activities.

### *Implications*

The results obtained here, combined with those of Schmit et al. (1995), have several implications. First, both studies suggest that adding a common frame-of-reference to personality items enables individuals to make more specific responses, thus allowing them to present themselves more accurately. This increased accuracy in self-presentation should lead to increased validity via reduced measurement error and increased reliability for scale scores composed of contextualized items. Indeed, although differences in alpha coefficients between contextualized and noncontextualized Conscientiousness scales appeared minor, alphas for contextualized Conscientiousness were significantly higher (see Feldt, 1980, for the significance test between dependent alpha coefficients) than alphas for noncontextualized Conscientiousness under both instructional conditions (general instructions,  $t(159) = 6.86, p < .01$ ; applicant instructions,  $t(140) = 2.00, p < .05$ ).

Thus, this research has significant implications for the creation of new personality measures. Some authors of personality tests have acted under the assumption that face valid items are easy to fake, and thus have attempted to develop items that are subtly related to the trait of interest. This practice, which follows from an adherence to socially desirable responding theory, has been rejected by the current research as well as by Schmit et al's (1995) study.

Clearly, writing personality items into specific contexts assists in obtaining accurate, relevant information about respondents, and improves personality test validity. This increase in personality test validity that results from increased item context-specificity should lead to more accurate selection decisions in organizational settings. Future research should be conducted in the field to examine this possibility.

From a theoretical standpoint, the current research raises questions about how to combine or integrate these new, context-specific personality items with traditional noncontextual personality items. This task may require the development of a situational taxonomy that can be integrated with more traditional, general taxonomies of personality to improve both the construct and predictive validity of personality theory and measurement.

Overall, these results demonstrate the importance of providing a common-frame-of-reference in personality items. The construction of context-specific personality items provides an opportunity to improve the prediction of performance and to add precision and depth to our understanding of personality.

References

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1-26.

Camara, W. J., & Echternacht, G. (2000). *The SAT I and High School Grades: Utility in Predicting Success in College*. New York: College Entrance Examination Board.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Costa, P. T., Jr., & McCrae, R. R. (1992). *NEO-PI-R professional manual*. Odessa, FL: Psychological Assessment Resources.

Costa, P. T., Jr., McCrae, R. R., & Dye, D. A. (1991). Facet scales for Agreeableness and Conscientiousness: A revision of the NEO Personality Inventory. *Personality and Individual Differences, 12*, 887-898.

Cronbach, L. J., & Gleser, G. C. (1957). *Psychological Tests and Personnel Decisions*. Urbana, IL: University of Illinois Press.

Day, D. V. & Silverman, S. B. (1989). Personality and job performance: Evidence of incremental validity. *Personnel Psychology, 42*, 25-36.

Dollinger, S. J. & Orf, L. A. (1991). Personality and performance in "personality": Conscientiousness and Openness. *Journal of Research in Personality, 25*, 276-284.

Feldt, L. S. (1980). A test of the hypothesis that Cronbach's alpha reliability coefficient is the same for two tests administered to the same sample. *Psychometrika, 45*, 99-105.

Hogan, R. T. (1991). Personality and personality measurement. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2<sup>nd</sup> ed., Vol. 2, 873-919). Palo Alto, CA: Consulting Psychologists Press.

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75*, 95-108.

Hunter, J. E. & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72-98.

Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The big five revisited. *Journal of Applied Psychology, 85*, 869-879.

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin, 107*, 34-47.

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). The role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*, 660-679.

Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical employees. *Journal of Applied Psychology, 65*, 373-406.

Schmidt, F. L., Gast-Rosenburg, I., & Hunter, J. E. (1980). Validity generalization results for computer programmers. *Journal of Applied Psychology, 65*, 643-661.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262-274.

Schmidt, F. L., Hunter, J. E., & Caplan, J. (1981). Validity generalization results for two job groups in the petroleum industry. *Journal of Applied Psychology, 66*, 261-273.

Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, S. L. (1995). Frame-of-reference effects on personality scores and criterion-related validity. *Journal of Applied Psychology, 80*, 607-620.

Sechrest, L. (1963). Incremental validity: A recommendation. *Educational and Psychological Measurement, 23*, 153-158.

Sibert, P. C., & Ayers, J. B. (1989). Relationships between scores from the ACT, PPST, NTE, ACT COMP, and GPA. *Educational and Psychological Measurement, 49*, 945-949.

Taylor, J. B., Carithers, M., & Coyne, L. (1976). MMPI performance, response set, and the self-concept hypothesis. *Journal of Consulting & Clinical Psychology, 44*, 351-362.

Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44*, 703-742.

Wolfe, R. N., & Johnson, S. D. (1995). Personality as a predictor of college performance. *Educational & Psychological Measurement, 55*, 177-185.

Wright, J. C., & Mischel, W. (1987). A conditional approach to dispositional constructs: The local predictability of social behavior. *Journal of Personality and Social Psychology, 53*, 1159-1177.

Table 1

*Means, Standard Deviations, and Alpha Coefficients for Study Variables by Instructional Condition*

<i>Scale</i>	General Instructions <sup>a</sup>			Applicant Instructions <sup>b</sup>		
	<i>M</i>	<i>SD</i>	$\alpha$	<i>M</i>	<i>SD</i>	$\alpha$
Conscientiousness (noncontextual)	2.39	.45	.92	2.48	.49	.93
Conscientiousness (school-specific)	2.40	.53	.95	2.52	.53	.94
ACT	22.55	4.08	---	22.73	4.45	---
GPA	3.14	.68	---	3.17	.79	---

<sup>a</sup> *n* ranges from 161 to 185, due to missing data.

<sup>b</sup> *n* ranges from 142 to 157, due to missing data.

Table 2

*Correlations Among Study Variables by Instructional Condition*

	1	2	3	4
1. Conscientiousness (noncontextual)	---	.90**	.14*	.42**
2. Conscientiousness (school-specific)	.89**	---	.16*	.51**
3. ACT	.25**	.23**	---	.51**
4. GPA	.39**	.46**	.57**	---

*Note.* Correlations for the general instructions condition are above the diagonal, and correlations for the applicant instructions condition are below the diagonal.

\* $p < .05$ ; \*\* $p < .01$

Table 3

*Standardized Regression Weights, R-Squared Change, and Total R-Squared for Incremental Validity Analyses by Instructional Condition*

	General Instructions			
	Step 1	Step 2	Step 3	Step 4
1. ACT	.51**	.52**	.47**	.45**
2. Order of Item-type		.08	.04	.07
3. Conscientiousness (noncontextual)			.35**	-.25
4. Conscientiousness (school-specific)				.65**
$\Delta R^2$	.26**	.01	.12**	.07**
Total $R^2$	.26**	.27**	.39**	.46**
	Applicant Instructions			
	Step 1	Step 2	Step 3	Step 4
1. ACT	.57**	.56**	.50**	.50**
2. Order of Item-type		.05	.02	-.01
3. Conscientiousness (noncontextual)			.26**	-.23
4. Conscientiousness (school-specific)				.54**
$\Delta R^2$	.32**	.00	.06**	.05**
Total $R^2$	.32**	.32**	.39**	.44**

\* $p < .05$ ; \*\* $p < .01$

Table 4

*Standardized Regression Weights, R-Squared Change, and Total R-Squared for Incremental Validity Analyses (Full Sample)*

	Step 1	Step 2	Step 3	Step 4	Step 5
1. ACT	.54**	.54**	.54**	.48**	.47**
2. Instructional Condition		.02	.01	-.02	-.03
3. Order of Item-type			.07	.03	.03
4. Conscientiousness (noncontextual)				.31**	-.22*
5. Conscientiousness (school-specific)					.59**
$\Delta R^2$	.29**	.00	.00	.09**	.06**
Total $R^2$	.29**	.29**	.30**	.39**	.45**

\* $p < .05$ ; \*\* $p < .01$