Industrial and Organizational Psychology, **2** (2009), 501–503.
Copyright © 2009 Society for Industrial and Organizational Psychology. 1754-9426/09

Content Validity Does Matter for the Criterion-Related Validity of Personality Tests

H. KRISTL DAVISON AND MARK N. BING
University of Mississippi

In Murphy's (2009) focal article, ''Content Validation Is Useful for Many Things, but Validity Isn't One of Them,'' he focuses primarily on tests of knowledge, skills, and abilities, and whether content validity leads to criterion-related validity for such characteristics. With respect to personality testing, however, he notes that the ''jury is still out over the question of whether content validity makes any real differences in contexts such as . . . personality tests'' (p. 462). However, we believe that research on the frame-of-reference effect in personality testing clearly indicates that increased content validity of personality tests enhances their criterion-related validity.

Several studies thus far have investigated the frame-of-reference effect in personality (e.g., Bing, Whanger, Davison, & VanHook, 2004; Lievens, DeCorte, & Schollaert, 2008; Schmit, Ryan, Stierwalt, & Powell, 1995). The fundamental idea underlying the frame-of-reference effect is that contextualizing self-report personality test items to specific situations (e.g., academic, work), and thus improving their content validity, can facilitate more accurate self-presentations on the part of test takers. Specifically, when items are *not* contextualized to a particular situation, then test takers may be unsure as to which ''self'' to present when responding to the item—the work self, the school self, the home self, the community self, and so forth. If the test taker is uncertain as to whether the item calls for presentation of the work self or the nonwork self, then accurate self-presentation may be hindered, and the criterion-related validity of the personality test would be reduced. In contrast, providing a specific situational context to the personality test item allows the test taker to respond to the item with the appropriate self, which increases the content validity of the test by definition, and then the subsequent criterion-related validity according to recent research (e.g., Bing et al., 2004; Schmit et al., 1995).

Take, for example, the following personality test item: ''I strive for excellence in everything I do.''

When responding to this item, different test takers may rely on different core self-evaluations, such that one test taker might consider the item in the context of her work activities, another might respond in the context of his school activities, and a third might respond in the context of her community activities. Thus, because different test takers are answering this item with their own imposed different contextual *frames of reference*, they are in essence answering

Correspondence concerning this article should be addressed to H. Kristl Davison.
E-mail: kdavison@bus.olemiss.edu
Address: Department of Management, School of Business Administration, University of Mississippi, University, MS 38677
H. Kristl Davison and Mark N. Bing, Department of Management, School of Business Administration, University of Mississippi

https://doi.org/10.1111/j.1754-9434.2009.01182.x Published online by Cambridge University Press

different personality test items. These different interpretations of the item by different test takers also mean that general, noncontextualized items will differ in their *psychological* content validity across test takers.

In contrast, consider the following personality test item: ''I strive for excellence in everything I do *at school*.''

By altering the content of the item such that it is contextualized to a school setting, the content validity of the item should thus increase at least with respect to the criterion of school performance. Owing to this contextualization of the item, now when test takers respond they will rely on a common self-evaluation, and thus they will have a common frame of reference, namely the school context. Consequently, all test takers will be answering the *same* personality test item given the enhanced, specific content of that item.

Research has shown that contextualizing personality test items to a relevant situation (e.g., work or school) increases the reliability and criterion-related validity of these tests. Using a student sample, Schmit et al. (1995) found that standardizing and providing a criterion-relevant frame-of-reference in conscientiousness test items reduced measurement error and increased both reliability and criterion-related validity in the prediction of college grade point average (GPA). Bing et al. (2004) also found that not only did contextualizing conscientiousness test items lead to increased reliability and criterion-related validity in the prediction of GPA but also the contextualized personality test had *incremental validity* over the noncontextualized personality test. This *incremental validity* was obtained despite the high level of multicollinearity between the contextualized and noncontextualized conscientiousness tests ($r = .89–.90$). Further, Bing et al. showed that this increase in criterion-related validity was present even after statistically controlling for the increased reliability. Thus, the increase in criterion-related validity was not simply due to greater reliability of measurement but was in fact a result of increased criterion-relevance of the personality test items achieved via the specified content of the items, and thus the improved content validity of the items.

More recently, Bing, Davison, and Whanger (2006) found further support for the necessity of contextualizing personality test items to a criterion-relevant context in order to increase criterion-related validity. Specifically, in a student sample, they contextualized conscientiousness test items to a school or community setting and collected both school-relevant criteria as well as community participation criteria. They found that community-contextualized conscientiousness predicted community participation, whereas school-contextualized conscientiousness did not. They also found that school-contextualized conscientiousness predicted class attendance, whereas community-contextualized conscientiousness did not. Again, the increased content validity of the items for particular contexts enhanced criterion-related validity for predicting criteria within those particular contexts.

Similarly, Lievens et al. (2008) found that use of a frame of reference reduced within-person inconsistency in a student sample and led to higher criterion-related validity in the prediction of GPA. Importantly, their study also showed that, consistent with Bing et al.'s (2004, 2006) findings, reducing inconsistency is not sufficient for increasing validity and that the items must be contextualized in a way that is relevant to the criterion of interest. Specifically, school-contextualized achievement, self-discipline, and conscientiousness tests were significantly better predictors of GPA than the same tests contextualized to a work setting. Thus, once more, it was shown that increased content validity (for a particular context) led to increased criterion-related validity.

Other studies (e.g., Hunthausen, Truxillo, Bauer, & Hammer, 2003), as well as the studies described here (Bing et al., 2004; Lievens et al., 2008; Schmit et al., 1995), have also used instruction-level manipulations to provide a frame-of-reference and have also generally found increased

https://doi.org/10.1111/j.1754-9434.2009.01182.x Published online by Cambridge University Press

criterion-related validity for these manipulations, although the frame-of-reference effect seems to be somewhat stronger when item-level manipulations are used. Overall, the results of these various studies indicate very clearly that enhancing content validity of personality test items via contextualizing them to a specific criterion-relevant situation enhances their criterion-related validity. Thus, it appears that *content validity is very relevant for increasing the criterion-related validity of personality tests*.

References

Bing, M. N., Davison, H. K., & Whanger, J. C. (2006). *Increases in validity as a function of frame-of-reference control in personality testing: Construct alteration vs. reliability enhancement*. In D. H. Kluemper (Chair), *Self-report Measurement Issues in Micro Organizational Research*. Symposium presented at the 66th Annual Conference of the Academy of Management, Atlanta, Georgia.

Bing, M. N., Whanger, J. C., Davison, H. K., & VanHook, J. B. (2004). Incremental validity of the frame-of-reference effect in personality scale scores: A replication and extension. *Journal of Applied Psychology*, *89*, 150–157.

Hunthausen, J. M., Truxillo, D. M., Bauer, T. B., & Hammer, B. L. (2003). A field study of frame-of-reference effects on personality test validity. *Journal of Applied Psychology*, *88*, 545–551.

Lievens, F., DeCorte, W., & Schollaert, E. (2008). A closer look at the frame-of-reference effect in personality scale scores and validity. *Journal of Applied Psychology*, *93*, 268–279.

Murphy, K. R. (2009). Content validation is useful for many things, but validity isn't one of them. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *2*, 453–464.

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.

https://doi.org/10.1111/j.1754-9434.2009.01182.x Published online by Cambridge University Press