Journal of Career Assessment

Development and Validation of the Calling and Vocation Questionnaire (CVQ) and Brief Calling Scale (BCS)

Bryan J. Dik, Brandy M. Eldridge, Michael F. Steger and Ryan D. Duffy

The online version of this article can be found at:
http://jca.sagepub.com/content/20/3/242

Published by:

$SAGE

http://www.sagepublications.com

Additional services and information for *Journal of Career Assessment* can be found at:

Email Alerts: http://jca.sagepub.com/cgi/alerts

Subscriptions: http://jca.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations: http://jca.sagepub.com/content/20/3/242.refs.html

>> Version of Record - Jul 11, 2012

OnlineFirst Version of Record - Jan 24, 2012

What is This?

Development and Validation of the Calling and Vocation Questionnaire (CVQ) and Brief Calling Scale (BCS)

Bryan J. Dik[1], Brandy M. Eldridge[1],
Michael F. Steger[1,2], and Ryan D. Duffy[3]

Abstract

Research on work as a calling is limited by measurement concerns. In response, the authors introduce the multidimensional Calling and Vocation Questionnaire (CVQ) and the Brief Calling scale (BCS), instruments assessing presence of, and search for, a calling. Study 1 describes CVQ development using exploratory and confirmatory factor analysis (CFA) in a cross-validated split-sample approach with 456 undergraduates. The CVQ contained six reliable subscales that form CVQ-Presence and CVQ-Search scores, which demonstrated moderate 1-month test–retest reliability, good fit to a six-factor structure, and initial support for construct validity. Study 2 reports a multitrait–multimethod analysis with 134 undergraduates and 365 informants. Self-reported CVQ and BCS scores moderately to strongly correlated with informant reports and scores for both instruments correlated in hypothesized directions with work hope, prosocial work motivation, life meaning, and the search for meaning. CVQ and BCS scores provide psychometrically sound measures of calling, with the CVQ offering the potential for more fine-grained, multidimensional analyses.

Keywords
calling, vocation, meaningful work, instrument development

The idea that work can be approached as a calling or vocation has a long history (Hardy, 1990), but a short past within the social sciences, which only recently have begun to investigate the construct empirically (e.g., Bunderson & Thompson, 2009; Davidson & Caddell, 1994; Duffy & Sedlacek, 2007; Elangovan, Pinder, & McLean, 2010; Serow, 1994; Steger, Pickering, Shin, & Dik, 2010; Wrzesniewski, McCauley, Rozin, & Schwartz, 1997). The notion of calling seems to lie at the heart of a holistic understanding of work in the context of life, highlighting concerns such as career

[1]Department of Psychology, Colorado State University, Fort Collins, CO, USA
[2]Department of Psychology, North-West University, Potchefstroom, South Africa
[3]Department of Psychology, University of Florida, Gainesville, FL, USA

Corresponding Author:
Bryan J. Dik, Department of Psychology, Colorado State University, Fort Collins, CO 80523, USA
Email: Bryan.Dik@ColoState.EDU

development, job satisfaction, well-being, and personal growth. Research suggests that it holds potential for uniting these and other domains (e.g., social justice, Dik, Duffy, & Steger, 2012; spirituality, Steger et al., 2010). However, as with many emerging areas of scholarship, measurement challenges have been a primary obstacle to overcome as research on calling proceeds (e.g., Dik & Duffy, 2009; Wrzesniewski, 2010). In response, the primary purpose of the current research was to develop and establish psychometric support for the Calling and Vocation Questionnaire (CVQ), a theoretically derived, multidimensional instrument designed to assess calling in the domain of work. A secondary purpose was to examine the construct validity of the Brief Calling scale (BCS), a highly efficient instrument designed to assess the extent to which people perceive that they have a calling, or are seeking one.

Research on Calling

As interest in the construct of calling has grown within psychology, sociology, and management, numerous studies have examined its importance to specific populations and how it relates to a host of other variables. Several have shown that a large proportion of college students and adults endorse the construct. Using data from over 5,000 college students, Duffy and Sedlacek (2010) found that more than 40% state that having a calling to a particular line of work was mostly or totally true of them; no substantive differences were found across gender or ethnicity. Similar percentages of working adults also view their careers as a calling (Bunderson & Thompson, 2009; Davidson & Caddell, 1994; Wrzesniewski et al., 1997), suggesting that the term is salient to a wide range of people.

Most research on calling has investigated its relation to career-related criterion variables. For college students, the presence of a calling has correlated positively with career decidedness, choice comfort, vocational self-clarity, career decision self-efficacy, work-related outcome expectations, academic satisfaction, and intrinsic work motivation (Dik, Sargent, & Steger, 2008; Duffy, Allan, & Dik, 2011; Duffy & Sedlacek, 2007; Steger et al., 2010). In other words, students more likely to endorse a calling tend to be more firm, comfortable, and confident in their career choices, view their future careers more positively, and are more satisfied in the academic domain. Similarly, calling has related positively with beneficial work-related variables among employed adults. For example, those with a calling tend to display greater occupational identification, occupational importance, career commitment, work meaningfulness, perceived organizational duty, work zest, and work satisfaction (e.g., Bunderson & Thompson, 2009; Davidson & Caddell, 1994; Duffy, Dik, & Steger, 2011; Peterson, Park, Hall, & Seligman, 2009; Serow, 1994; Steger & Dik, 2009; Wrzesniewski et al., 1997). To a somewhat lesser degree, calling also has been positively associated with well-being. College students endorsing a calling tend to be slightly more satisfied with their lives and report higher levels of life meaning (Duffy & Sedlacek, 2010; Steger & Dik, 2009; Steger et al., 2010); these linkages are somewhat stronger when assessed with employed adults (Bunderson & Thompson, 2009; Peterson et al., 2009). To summarize, endorsing higher levels of a calling appears to be related to both positive work-related and general well-being criterion variables.

Measurement Issues

When discussing research on calling, it is critical to point out concerns with how the construct has been measured. Calling measures have been constructed almost exclusively on an ad hoc basis according to the needs of individual studies. Often, they operationalize a construct that is only loosely defined in the studies from which measures have originated, and thorough psychometric evidence is rarely reported. Typically, measurement strategies have consisted of the following: (a) categorizing participants into "calling," "job," or "career" orientations based on ratings of vignettes describing different

work orientations (e.g., Peterson et al., 2009; Wrzesniewski et al., 1997); (b) scales developed for specific populations of workers (e.g., Bunderson & Thompson, 2009; Dobrow & Tosti-Kharas, 2011; Serow, 1994); and (c) multiitem unidimensional scales (e.g., Dobrow & Tosti-Kharas, 2011; Dreher, Holloway, & Shoenfelder, 2009; Treadgold, 1999). Each type of instrument carries limitations. Categorization techniques have reliability concerns and may limit the capacity to assess the full range of variability that likely exists among the endorsement of a calling. Instruments developed for specific occupations are often useful in their immediate context, but the extent to which such instruments can be used with other samples is unclear. Finally, unidimensional strategies are unable to examine complexities that may exist in how people express or experience a sense of calling, given that the construct has been conceptualized as multidimensional in nature (e.g., Dik & Duffy, 2009; Elangovan et al., 2010).

The Current Research

Without valid, theoretically grounded instruments to assess calling, it is more difficult for this line of research to advance in a rigorous and conceptually clear manner. Efforts to construct a new measure should begin with a clear definition of the construct being assessed (Dawis, 2000). We conceptualize calling in line with the three-part definition proposed by Dik and Duffy (2009), who described a calling as "(1) a transcendent summons, experienced as originating beyond the self, to (2) approach a particular life role in a manner oriented toward demonstrating or deriving a sense of purpose or meaningfulness and (3) that holds other-oriented values and goals as primary sources of motivation" (p. 427, numbers added). Calling often is used interchangeably with the closely related term *vocation*. However, following authors who have noted changes in how the latter term has been used over time, Dik and Duffy defined vocation as an approach to a life role consisting of derivation of purpose or meaningfulness and other-oriented values or goals, but without the additional component of a perceived summons from an external source (i.e., vocation consists the second and third dimensions of calling). This manner of conceptualizing calling also posited two overarching and overlapping aspects of the construct: presence and search. That is, some individuals may perceive that they currently have a calling (presence), whereas others may not currently sense a calling but report that they are actively seeking one (search).

Assessing the "presence of" and "search for" (a) transcendent summons, (b) purposeful work, and (c) prosocial orientation dimensions of calling provides two primary advantages not shared by other measurement approaches. First, linking the measurement instrument to a clearly articulated definition of calling provides a level of conceptual clarity not currently available among most measures of calling. Second, assessing the construct multidimensionally allows exploration of more sophisticated research questions than can be examined using coarse unidimensional measures of calling. To summarize, the current research seeks to address a critical gap in literature by developing and validating a multidimensional instrument to measure calling, using Dik and Duffy's tripartite definition as a theoretical guide.

The second study reported in this article also provides a psychometric evaluation of the BCS, which assesses the presence of calling and search for a calling using scales consisting of just 2 items each. The BCS does not overcome the third limitation described above; it distinguishes presence and search dimensions but uses a unidimensional strategy to do so, collapsing the three components of calling into general calling scales. However, some researchers may at times desire to assess the extent to which a person believes she has a calling or is seeking one generally, without further examining the underlying dimensions. In such circumstances, brief instruments are especially useful, provided they yield reliable and valid scores. The BCS has been used in a few studies (e.g., Dik et al., 2008; Duffy & Sedlacek, 2007; Steger et al., 2010) with promising psychometric support, but this study provides the first thorough examination of the construct validity of its scores.

Study 1

The purpose of study 1 was to develop, refine, and provide initial psychometric evaluation of the CVQ, which was accomplished via the following steps: (a) development of an item pool, (b) a test of the items and establishment of an initial factor structure using exploratory factor analysis (EFA), (c) further refinement of the item pool using both EFA and CFA, (d) cross-validation of the measurement model using CFA, and (e) establishment of initial evidence for reliability and validity of scale scores. The first step was writing items, following available guidelines (e.g., Clark & Watson, 1995; Dawis, 2000), designed to assess both "presence" and "search" dimensions across each of the three domains specified in Dik and Duffy's (2009) definition of calling. Attempts were made to write items that were straightforward, that avoided colloquialisms, and that would be appropriate across a wide range of cultural groups. Attempts also were made to write a large pool of items that liberally oversampled the construct in ways that accounted for the disparate approaches to defining it, with the intent of allowing the data to "speak for themselves" when factor analytic strategies were applied. For example, because some definitions in the literature (e.g., Dalton, 2001) postulate attributions of calling to particular sources (e.g., God, social needs, one's own abilities), items were written that included these sources as well as items that were source-free. Similarly, because some definitions explicitly frame calling in a religious context (e.g., Davidson & Caddell, 1994), items were written to include the religious content contained in those definitions. Finally, some items were written which we expected could load negatively on factors assessing perceptions of work as a calling. This strategy followed directives by Clark and Watson (1995), who suggested that "the initial item pool (a) should be broader and more comprehensive than one's own theoretical view of the target construct and (b) should include content that ultimately will be shown to be tangential or even unrelated to core constructs" (p. 311).

This item generation process resulted in an initial pool of 180 items. A four-level response scale (1 = *not at all true of me*, 2 = *somewhat true of me*, 3 = *mostly true of me*, 4 = *absolutely true of me*) was selected to eliminate a neutral or midpoint response option (and therefore to minimize the central tendency response bias; Dawis, 2000) and to allow participants to classify themselves meaningfully without forcing unrealistically subtle distinctions between response levels. The 180 items were screened for clarity by a counseling psychologist (not otherwise involved in this study) with expertise on the topic of calling as well as scale construction, and by a group of undergraduate psychology volunteers. Their ratings and feedback prompted a round of revision, with some items rewritten and others deleted due to lack of clarity or redundancy, resulting in a revised pool of 156 items. Readability statistics available in Microsoft Word revealed that the items had a Flesch Reading Ease rating of 64.7 and a Flech–Kincaid Grade Level rating of 6.0, indicating a standard level of difficulty corresponding to approximately a sixth-grade reading level (Harvey, 1997). This revised item pool was subsequently administered to a sample of college students to facilitate further refinement and to provide initial evidence of reliability and validity.

Regarding validity evidence for the completed instrument, we hypothesized that those attending religious services more frequently and those indicating in absolute (i.e., yes or no) terms that calling is relevant to their approach to work would score higher on presence of calling and search for calling scales than those attending less often or reporting that calling is not relevant. We also examined differences across sex, race/ethnicity, and year in school, approaching these comparisons as research questions rather than as tests of directional hypotheses relevant for evaluating criterion-related validity. Finally, to establish initial evidence of construct validity, we administered measures of criterion variables with varying degrees of conceptual relatedness to calling. Following Dik and Duffy's (2009) suggestion that calling is an ongoing process rather than something to be discovered once and for all, we expected high correlations between scores reflecting the presence of, and search for, a calling. We also expected these correlations would be stronger than those of either calling variable

with intrinsic work motivation, which would in turn be stronger than correlations between calling variables and career decision self-efficacy, extrinsic work motivation, or life satisfaction.

Method

Participants and Procedure

A total of 456 participants were recruited from introductory psychology courses at a large, Western, public research university ($n = 360$) and two small, Midwestern, Christian liberal arts colleges ($n = 96$). College students were selected because of the developmental relevance of seeking and experiencing a calling for many students (e.g., Hunter, Dik, & Banning, 2010). The public research university sample was supplemented by students at the religiously affiliated institutions following a purposive sampling strategy guided by the expectation that calling may be particularly salient for participants with faith commitments. Participants reported a mean age of 18.78 ($SD = 1.42$, range of 17–31) and self-identified mostly as women (72.7%) and White/European American (85.1%; 2.4% = Latino/Hispanic American, 2.2% = Asian American and Pacific Islander; 1.3% = Black/African American; 0.9% = American Indian/Native American; 2.6% = biracial or multiracial; 1.8% = other). Although the sample included a high percentage of White women, Worthington and Whittaker (2006) noted that "representativeness in scale development research does not follow conventional wisdom—that is, it is not necessary to closely represent any clearly identified population as long as those who would score high and those who would score low are well represented" (p. 816).

Participants were offered partial course credit for participation. Students providing their informed consent were directed to an Internet survey that included the item pool and a brief demographic questionnaire. One month after the initial questionnaire (Time 1) was administered, an e-mail was sent to the public university students with a link to a second questionnaire (Time 2) containing the item pool and additional instruments (see below) for purposes of assessing test–retest reliability and establishing initial evidence of construct validity. A total of 333 (92.5%) of the original 360 public university participants participated at Time 2 (76% women, 83% White).

Instruments

In addition to the item pool, the following measures were administered at Time 2 for purposes of establishing initial evidence of convergent and discriminant validity. (All internal consistency reliability coefficients for the present study are reported in Table 1.)

Relevance of calling. The following items assessed whether participants identified with the term "calling": "As you think about your career, do you consider yourself as having a calling?" "Are you looking for a calling in your career?" and "Is the term 'calling' relevant as you think about your career?" Participants responded "yes" or "no" to each item.

Career decision self-efficacy. The short form of the Career Decision Self-Efficacy scale (CDSE-SF; Betz, Hammond, & Multon, 2005) was used to assess the extent to which participants felt confident in their ability to effectively navigate the career decision-making process. The CDSE-SF consists of 25 items that ask participants to rate their self-efficacy relative to specific decision-making tasks on a 5-point scale ($1 = no\ confidence\ at\ all$, $5 = complete\ confidence$). Higher scores reflect greater levels of career decision self-efficacy. Internal consistency reliability for the scale was reported by Betz, Hammond, and Multon (2005) to range from $\alpha = .94$ to $\alpha = .95$ with support for construct validity reflected in positive correlations with career decidedness, self-clarity, decisiveness, and occupational/educational knowledge.

Table 1. Means, Standard Deviations, and Intercorrelations of Study 1 Variables at Time 2

	1	2	3	4	5	6
1. CVQ presence	(.90)					
2. CVQ search	.77**	(.91)				
3. Career decision self-efficacy	.17**	.11	(.95)			
4. Intrinsic work motivation	.27**	.31**	.31**	(.79)		
5. Extrinsic work motivation	.12*	.16**	.11	.05	(.71)	
6. Life satisfaction	.14*	.09	.33**	.14*	−.01	(.88)
M	27.93	28.15	94.57	41.69	38.27	25.99
SD	8.23	8.06	15.20	5.87	5.50	5.65

Note. () = coefficient α; CVQ = Calling and vocation questionnaire.
*$p < .05$. **$p < .01$.

Intrinsic and extrinsic work motivation. The Work Preference Inventory (WPI; Amabile, Hill, Hennessey, & Tighe, 1994) consists of 30 items that use a 4-point scale (1 = *never or almost never true of me*, 4 = *always or almost always true of me*) to measure intrinsic and extrinsic work motivation. According to Amabile, Hill, Hennessey, and Tighe (1994), factor analyses support the structural validity of the scale scores, which also correlate in expected directions with other motivation measures, Myers–Briggs Type Indicator scores, cognitive curiosity, and Strong Interest Inventory scores. Internal consistency coefficients were initially reported to be α = .82 for the intrinsic scale scores and α = .76 for extrinsic scale scores.

Life satisfaction. The Satisfaction with Life scale (SWLS; Diener, Emmons, Larson & Griffin, 1985) is a frequently used 5-item scale assessing participants' cognitive appraisal of well-being (e.g., In most ways my life is close to the ideal). Items use a 7-level response scale (1 = *strongly disagree*, 7 = *strongly agree*); higher scores indicate greater satisfaction with life. Evidence of reliability and validity for SWLS is well established (e.g., Pavot & Diener, 1993).

Results

Refining the Item Pool

The first step in our analytic strategy (see Steger, 2006) was to use EFA coupled with parallel analysis to (a) verify the presence of factors specified by Dik and Duffy's (2009) definition in the item pool, and (b) identify extraneous item content. EFA was used to verify that the desired item content remained intact through efforts to reduce the size of the item pool. To maintain an adequate item-to-participant ratio at this stage (Kahn, 2006), we conducted the initial EFAs on the full sample of 456 participants. Anticipating that factors would be intercorrelated, we subjected the correlation matrix for the 156 items to a principle axis factoring (PAF) analysis with oblique, promax rotation (see Kahn, 2006; Steger, 2006; Worthington & Whittaker, 2006). The initial solution demonstrated that the matrix was factorable (Kaiser–Meyer–Olkin measure of sampling adequacy = .95, Bartlett's Test of Sphericity − χ^2(12,090, N = 456) = 67,231.46, p < .001, Determinant = 1.46E − 075). Parallel analysis (Hayton, Allen, & Scarpello, 2004) indicated that nine factors should be extracted. These nine factors explained 60.42% of the variance. An examination of item content across these factors revealed that five factors corresponded to the six dimensions specified in Dik and Duffy's (2009) definition: Transcendent summons—presence (20 items that left open the source of the summons); Transcendent summons—search (11 source-free items); Purposeful work—presence (15 items); Purposeful work—search (16 items); and one factor containing presence and search items related to prosocial motives (29 items). The remaining factors consisted of source-specific

items (religion, spirituality, 28 items, or fate as source, 3 items), "anticalling" items (materialistic values, lack of purpose, lack of prosocial motives, 26 items), and a factor consisting of items referencing a need for financial security and a need to follow a family legacy (8 items). These latter four factors were deemed outside of the content area specified by Dik and Duffy's (2009) definition, and items that loaded on these factors were deleted, resulting in a final pool of 91 items. A second EFA using PAF with promax rotation confirmed that deleting these items left the five definition-relevant factors intact.

The next step aimed to accomplish two objectives: (a) shorten the measure to increase its utility for research purposes while maintaining high reliability and (b) equalize the number of items on each scale, since there is no theoretically indicated reason for differentially weighting the dimensions. Because the item pool was smaller, we were able to use a split-sample method, employing an iterative process with multiple EFAs (PAF, promax) on a randomly selected portion of the sample, representing roughly 50% of completed surveys ($n = 234$). This approach allows for cross-validation of the final factor structure in a subsample that is relatively independent from efforts to refine the item pool. This reduces capitalizing on sample-specific variance. In this case, the full sample was used to eliminate a number of items, but the split-sample method protected half the sample from decisions about eliminating items that contain the specific content identified by Dik and Duffy (2009), meaning this half of the sample provides a relatively "clean" estimate of the conformance of the final item pool to the theoretically suggested factor structure.

First, the 11 items with factor loadings less than .60 on their intended factor were eliminated (see Watson, Clark, & Tellegen, 1988). Additional item elimination decisions were based on redundancy of items, adherence to Dik and Duffy's (2009) conceptualization of calling dimensions, and α values if particular items were deleted. The item deletion process ceased when αs for scale scores dipped below .85 or scale brevity resulted in inadequate coverage of the definition of the construct. These procedures resulted in a revised item pool of 24 items, all of which were positively scored. These remaining 24 items were subjected to a final PAF with promax rotation, with six factors extracted. Evidence suggested a factorable matrix (Kaiser–Meyer–Olkin measure of sampling adequacy = .88, Bartlett's Test of Sphericity $- \chi^2(276, N = 230) = 3{,}629.70, p < .001$, Determinant = 4.72E − 008). The prosocial orientation items previously loading on a single factor split into separate Presence-Prosocial Orientation and Search-Prosocial Orientation factors, and all other items also loaded on their expected factors. The eigenvalues for the six factors were 8.06, 3.78, 2.58, 1.71, 1.13, and 0.76, respectively, accounting for 33.56%, 15.76%, 10.74%, 7.13%, 4.70%, and 3.18% of the variance, for a total of 75.09%.

Final revisions to the item pool. Factor structure is not the only criteria for judging the items of a new measure. Efforts also should be made to avoid introducing method bias in an item pool. Using EFA to refine an item pool can maximize item homogeneity, which may extend to item valence. As noted above, all of the items that made it through the EFA analyses were positively scored. To reduce bias (e.g., acquiescence) that may arise from having all positively scored items, we reintroduced eight reverse-scored items that were consistent with adherence to Dik and Duffy's (2009) calling dimensions, but which were eliminated earlier. Thus, a total of 32 items were passed through to finalizing and testing the measurement model.

Finalizing and Testing the Measurement Model

Although EFA provides valuable information about the patterns of interrelations among items, CFA is the preferred method for finalizing psychological measures and evaluating the fit of items to theoretical measurement models. The major advantage of CFA over EFA approaches to scale construction is that CFA provides empirical indices of model fit (so models can be compared) and

modification indices (so the model can be improved to maximize fit; see Floyd & Widaman, 1995). CFA quantifies the discrepancy between hypothesized relations among variables (e.g., an item is correlated with one subscale's latent factor and not with another) and the relations actually observed in the data. We used the structural equation modeling software AMOS 6.0 (Arbuckle, 2005) to evaluate how well the hypothesized six-factor structure fit the relations among the 32 remaining items in the item pool (referred to as Model 1), initially using the same first half of the split sample. When necessary, the results of CFA were used to eliminate items according to these three criteria: (a) poor model fit, (b) standardized factor loadings less than .60, and (c) summed modification indices greater than 25.00.

Model 1 did not provide a good fit to the data, according to Hu and Bentler's (1999) criteria of Confirmatory Fit Index (CFI) and Nonnormed Fit Index (NNFI) close to or greater than .95 and standardized root mean squared residual (SRMR) and root mean square approximation of error (RMSEA) close to or less than .05, $\chi^2(df = 449) = 924.88$, $p < .001$; CFI = .88; NNFI = .87; SRMR = .08; RMSEA = .07; RMSEA 90% confidence interval (CI) = [.06, .10]. According to our criteria, we eliminated 8 items, most of them negatively scored. We next tested Model 2, consisting of 24 items, with each subscale consisting of a latent factor that loaded onto 4 items.

Model 2 provided acceptable fit to the data, $\chi^2(df = 237) = 410.87$, $p < .001$; CFI = .94; NNFI = .94; SRMR = .05; RMSEA = .06; RMSEA 90% CI [.05, .07]. Freeing four pairs of item error estimates would have improved the model above all of Hu and Bentler's (1999) criteria (i.e., Model 3), $\chi^2(df = 233) = 364.63$, $p < .001$; CFI = .96; NNFI = .95; SRMR = .05; RMSEA = .05; RMSEA 90% CI [.05, .06]. In the interests of parsimony and in an effort to avoid exploiting sample-specific error, we chose to accept Model 2 as the final CVQ solution. Table 2 presents the factors and standardized regression weights (for both halves of the sample) for this solution.

Because CFA was used to make changes to the original model tested using the first half of the split sample, validation of the final model using the second half of the data set was attempted. The fit was actually slightly better in the cross-validation half of the data set, $\chi^2(df = 237) = 391.29$, $p < .001$; CFI = .96; NNFI = .95; SRMR = .04; RMSEA = .06; RMSEA 90% CI [.05, .07]. The fit of this model was compared to two alternative models using the Akaike Information Criteria (AIC) statistic, which is appropriate for nonnested models. The AIC of the final model identified above was 517.29. The AIC of the first alternative model, in which one latent factor loaded on all "search" items and a second latent factor loaded on all "presence" items, was 2,029.26; the AIC of an alternative model in which all items loaded on a single factor was 2,193.61. Because smaller values indicate a better fit, it can be concluded that the six-factor CVQ model fit better than plausible alternative models. Thus, we concluded that the solution in Table 3 provided a good measure with good initial evidence of structural validity. The final CVQ therefore consisted of 24 items, comprised of three 4-item subscales (Presence-Transcendent Summons, Presence-Purposeful Work, Presence-Prosocial Orientation) that reflect the presence of calling (CVQ-Presence) and three 4-item subscales (Search-Transcendent Summons, Search-Purposeful Work, Search-Prosocial Orientation) that reflect the search for calling (CVQ-Search). Combining the respective Presence and Search subscales into summed Presence and Search total scores reduced the fit of the model. It appeared that this decrease in model fit was primarily caused by the amount of variance shared by the presence and search Purposeful Work subscales and the presence and search Prosocial Orientation subscales. This shared variance can be seen in the high positive correlations between these subscale pairs on Table 2. However, we feel that the advantages of creating Presence and Search total scores in terms of increased usefulness of the CVQ outweighs the better structural fit achieved by keeping each subscale separate. Thus, wherein CFA suggests six relatively independent subscales, we provide analyses based on CVQ-Presence and CVQ-Search scores as well.

Table 2. Standardized Regression Weights for CVQ Items Computed on First and Second Split-Sample Halves

Item	Transcendent Summons-Presence		Transcendent Summons-Search		Purposeful Work-Presence		Purposeful Work-Search		Prosocial Orientation-Presence		Prosocial Orientation-Search	
	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd
1	.89	.86										
11	.75	.76										
23	.84	.91										
8(r)	−.62	−.54										
2			.93	.88								
19			.61	.59								
13			.80	.84								
18			.81	.78								
3					.84	.87						
15					.81	.84						
33					.73	.82						
20					.78	.81						
4							.88	.92				
6							.76	.82				
14							.73	.77				
21							.74	.86				
22									.75	.77		
12									.85	.90		
9									.86	.91		
17									.64	.74		
16											.81	.87
10											.85	.87
7											.86	.91
5											.89	.87

Note. CVQ = Calling and vocation questionnaire; (r) = reverse scored. For each CVQ dimension, weights in the first and second columns were computed using the first and second halves of the split-half sample, respectively. All coefficients are significant at $p < .001$.

Table 3. Estimates of Relations Among CVQ Factor Scores From Confirmatory Factor Analyses

	1	2	3	4	5	6
1 Presence—Transcendent Summons		.24***	.58***	.40***	.27***	.24***
2 Search—Transcendent Summons	.36***		.25***	.46***	−.08	.03
3 Presence—Purposeful work	.66***	.33***		.71***	.46***	.48***
4 Search—Purposeful work	.54***	.50***	.79***		.34***	.47***
5 Presence—Prosocial orientation	.42***	.12	.59***	.47***		.78***
6 Search—Prosocial orientation	.29***	.12	.47***	.52***	.84***	

Note. Coefficients from the first split-half sample are above the diagonal; Coefficients from the second split-half sample are below the diagonal.
***$p < .001$.

Descriptive Statistics

Means, standard deviations, and intercorrelations for continuous variables at Time 2 are reported in Table 1. Roughly half of participants identified with the notion of calling, with 52.3% responding *yes* to the item, "As you think about your career, do you consider yourself as having a calling?"

Slightly higher percentages responded *yes* to "Are you looking for a calling in your career?" (53.2%) and "Is the term 'calling' relevant as you think about your career?" (54.3%).

Relations With Demographics

Mean comparisons at Time 2 evaluated the effect of demographic variables on CVQ scores. For CVQ-Presence, women ($M = 28.77$, $SD = 8.28$) scored significantly higher than men ($M = 25.14$, $SD = 7.16$), $t(328) = 3.419$, $p = .001$. Women ($M = 28.93$, $SD = 8.19$) also scored significantly higher than men ($M = 25.72$, $SD = 6.80$) on CVQ-Search, $t(326) = 3.083$, $p = .002$. No differences were found across ethnicity categories for CVQ-Presence, $F(6.323) = .838$, $MSE = 56.149$, $p = .541$ or CVQ-Search, $F(6, 321) = 1.559$, $MSE = 98.724$, $p = .159$. Similarly, no differences were found for year in school on either CVQ-Presence, $F(3, 316) = 1.157$, $MSE = 77.896$, $p = .326$, or CVQ-Search, $F(3, 315) = .307$, $MSE = 19.647$, $p = .827$.

As hypothesized, there was a significant effect for attendance at religious services for CVQ-Presence, $F(5, 323) = 8.204$, $MSE = 495.219$, $p = .000$, and CVQ-Search, $F(5, 321) = 9.730$, $MSE = 550.329$, $p = .000$. Post hoc Tukey's HSD tests showed that participants who report attending religious services more than once a week had significantly higher scores on CVQ-Presence and CVQ-Search than those who reported attending services two to six times a year, one to three times a year, or never. Participants who reported attending services one to three times a month also had significantly higher scores for both presence of and search for calling than those who indicated they never attend. All other comparisons yielded nonsignificant differences. Also as hypothesized, participants who responded *yes* to any of the three relevance of calling items scored significantly higher than those who responded *no* to all 3 items, both for CVQ-Presence (*yes*: M = 30.95, SD = 7.65; *no*: M = 22.53, SD = 6.24), $t(333) = 10.308$, $p = .000$, and CVQ-Search (*yes*: M = 31.06, SD = 7.30; *no*: M = 22.99, SD = 6.67), $t(331) = 9.988$, $p = .000$.

Internal Consistency and Test–Retest Reliability

At Time 1, internal consistency reliability coefficients for CVQ subscales were as follows: Presence-Transcendent Summons, $\alpha = .85$; Search-Transcendent Summons, $\alpha = .86$; Presence-Purposeful Work, $\alpha = .88$; Search-Purposeful Work, $\alpha = .88$; Presence-Prosocial Orientation, $\alpha = .88$; and Search-Prosocial Orientation, $\alpha = .92$. Total scores revealed similar internal consistency, with $\alpha = .89$ for CVQ-Presence and $\alpha = .87$ for CVQ-Search. Similar values also were found at Time 2, with a range of $\alpha = .83$ to $\alpha = .93$ for CVQ subscales (see Table 1 for total score values). Reliability of CVQ scores was further evaluated by correlating scores across the 1 month interval between Time 1 and Time 2. These test–retest coefficients for total scores were $r = .75$ for CVQ-Presence and $r = .67$ for CVQ-Search. Test–retest coefficients for subscale scores were as follows: Presence-Transcendent Summons, $r = .67$; Search-Transcendent Summons, $r = .62$; Presence-Purposeful Work, $r = .63$; Search-Purposeful Work, $r = .60$; Presence-Prosocial Orientation, $r = .66$; Search-Prosocial Orientation, $r = .67$.

Convergent and Discriminant Validity

Correlations among all continuous variables at Time 2 are reported in Table 1. As hypothesized, the most robust relationship existed between presence of and search for calling. Among remaining variables, CVQ-Presence was most strongly related to intrinsic work motivation, with weaker but still positive correlations found for CVQ-Presence with CDSE, life satisfaction, and extrinsic work motivation. CVQ-Search was moderately correlated with intrinsic motivation and positively but weakly

correlated with extrinsic work motivation; its relations with CDSE and life satisfaction were weak and not significant. This pattern of relationships supported hypotheses.

Summary and Discussion

Study 1 reported on the use of EFA and CFA to establish empirical support for the CVQ, a multidimensional measurement instrument designed to assess the facets of Dik and Duffy's (2009) conceptualization of calling. Results using a split-sample, cross-validation approach to refining the item pool and testing the measurement model provide support for a 24-item CVQ with strong evidence of internal consistency reliability. Test–retest reliability coefficients suggested that scores on the constructs are moderately stable over a 1-month period, which is reasonable given that calling is conceptualized as a generally stable yet malleable construct (e.g., Dik, Duffy, & Eldridge, 2009). Results indicated that women scored slightly higher than men on both presence and search dimensions. Scores also were higher for those reporting more frequent (compared to less frequent) attendance at religious services, and for those indicating that calling is a relevant concern in their career development, providing initial support for criterion-related validity of CVQ scores. No other significant demographic differences were found. Initial evidence was also found supporting the convergent and discriminant validity of scale scores, given that scores on the calling scales correlated more highly with each other than with less conceptually similar criterion variables. The CVQ is unique among instruments designed to measure calling in that it is grounded in a clearly defined conceptualization of calling and provides a multidimensional approach to assessing the construct. However, the initial validity evidence in Study 1 is limited to the criterion variables included in the study. Furthermore, although validity evidence was examined using data collected at a later point in time, participants were the same as those providing the data used to develop and refine the instrument. Therefore, a more comprehensive and formal test of the construct validity of CVQ-Presence and CVQ-Search scores was desired, using an independent sample and a wider array of criterion variables.

Study 2

The purpose of this study was to further evaluate the convergent and discriminant validity of both CVQ and BCS scores using a multitrait–multimethod matrix design. This design provides a comprehensive means of investigating construct validity by testing how well scores on a measure succeed in assessing their targeted construct without measuring other constructs (Campbell & Fiske, 1959). Specifically, evidence for convergent validity is established when scores on measures of the same trait are significantly correlated, whether they are measured using the same method (monotrait–monomethod) or different methods (monotrait–heteromethod). Evidence for discriminant validity is established when convergent correlations (i.e., between measures of the same trait) are higher than correlations with scores on measures of conceptually dissimilar traits, whether they are measured using the same method (heterotrait–monomethod) or different methods (heterotrait–heteromethod). The two methods employed in this study were self-report and informant-report. Conceptually, if the CVQ and BCS meet the measurement claims we make for them, scores on the instrument should reflect not just self-perception, but a pattern of behavior observable by others. Hence, patterns of convergent and discriminant correlations should be similar for self-report and informant-report methods. To keep the number of comparisons manageable, we focused on total scores (i.e., CVQ and BCS Presence and Search) in this study rather than also examining CVQ subscale scores.

Because convergent validity is established when different measures of the same trait are highly correlated, we expected significant and very strong positive correlations for CVQ-Presence with BCS-Presence scores, and for CVQ-Search with BCS-Search scores. We also expected significant

positive correlations between presence scale scores on both instruments with scores on the Wrzesniewski, McCauley, Rozin, and Schwartz (1997) calling measure, although we expected that differences in definitions of calling would result in lower correlations involving this latter scale than those found between the CVQ and BCS scales. We also expected significant positive correlations between self- and informant-reports on the same variables.

In contrast, because discriminant validity is established when convergent correlations are stronger than correlations of dissimilar variables, we expected that correlations among calling variables (especially those between like CVQ and BCS scales) would be stronger than correlations between CVQ or BCS scores and scores on measures of less conceptually similar variables. In this study, these other variables consist of work hope, prosocial work motivation, meaning in life, and life satisfaction; each of these are expected to be positively related to calling, but since they are measures of constructs other than calling, these correlations should be weaker in magnitude than correlations of the different measures of calling with each other. Also, we expected different methods of measuring the same variables (monotrait–heteromethod) to be more highly correlated than the same method of measuring different variables (heterotrait–monomethod). Finally, this study compared the patterns of convergent and discriminant correlations of CVQ, BCS, and Wrzesniewski et al.'s single-item scores to assess the extent to which the three scales are differentially supported by construct validity evidence.

Method

Participants and Procedure

Participants providing self-report scores were drawn from psychology courses at a large research university in the Western United States and compensated with course credit. They were asked to send to the researchers names and e-mail addresses of two to four people who knew them well (such as friends, parents, or siblings) and who would be willing to respond to a brief survey. Informants were then e-mailed a link to the same items to which referring participants responded, and were instructed to rate each item as if they were the referring participant.

A total of 134 participants referred two or more informants who provided usable data. These referring participants were predominantly female (81.3%) and White (84.2%; 5.8% self-identified as multiracial, 2.9% as Latina/Latino, 1.4% as Asian or Pacific Islander, .7% as African American, 1.4% as other, and 3.6% did not respond) and reported a mean age of 19.33 (*SD* = 3.10, range of 17–49). The 365 informants also were predominantly female (65.8%) and White (83.8%; 4.4% self-identified as Latina/Latino, 3.6% as multiracial, 2.7% as African American, 1.4% as American Indian or Alaskan Native, 1.4% as Asian or Pacific Islander, 1.6% as other, and 3.6% did not respond) and reported a mean age of 32.03 (*SD* = 15.57, range of 13–81).

Scores provided by the informants for each referring participant were averaged to form a composite ''informant'' score for that participant; these were used to compute correlations between self and informant on each pairing of scale scores. Referring participants had an average of 2.72 informants providing usable data.

Instruments

The 24 CVQ items and the 5-item SWLS used in Study 1 also were used in Study 2. Other measures included the following:

BCS. The BCS is a 4-item scale providing unidimensional scores that assess the presence of, and search for, a calling. The BCS defines calling as "a person's belief that she or he is called upon (by the needs of society, one's own inner potential, by God, by a Higher Power, etc.) to do a particular

kind of work.'' Items use a five-level response scale and are provided in Appendix B. The 2 items in the BCS-presence scale and the 2 items in the BCS-search scale were reported to correlate $r = .81$ and $r = .75$, respectively (Duffy & Sedlacek, 2007). Scores on both scales have correlated in expected directions with criterion variables such as sense of calling (assessed using a career development strivings strategy; see Dik et al., 2008), career decidedness, self-clarity, decision self-efficacy, meaning in life, intrinsic work motivation, and materialism (Dik et al., 2008; Duffy & Sedlacek, 2007).

Wrzesniewski et al.'s (1997) calling paragraph. Wrzesniewski et al. operationalized Bellah, Madson, Sullivan, Swidler, and Tipton (1986) tripartite distinction of Job, Career, and Calling orientations to work by constructing paragraphs providing a description of each orientation. This conceptualization of calling differed from Dik and Duffy's (2009) definition in that it reflects an approach to work characterized by a sense of social value and personal fulfillment, and did not include a transcendent summons component. In the current study, the calling paragraph was administered along with Wrzesniewski et al.'s (1997) single-item scale for rating the extent to which the person depicted in the paragraph is $0 = $ *not at all like me*, $1 = $ *a little [like me]*, $2 = $ *somewhat [like me]*, or $3 = $ *very much [like me]*. Wrzesniewski et al. reported that responses to the calling item correlated in predicted directions with responses to 18 true–false items also designed to differentiate Job, Career, and Calling orientations.

Work hope. The Work Hope scale (WHS; Juntunen & Wettersten, 2006) is a 24-item measure designed to assess work hope, a motivational state, with positive valence, characterized by having agency and pathways for pursing work-related goals held by the individual. Items (e.g., I expect to do what I really want to do at work) use a 7-point response scale ($1 = $ *strongly disagree*, $7 = $ *strongly agree*). Scores on the WHS have demonstrated high internal consistency ($\alpha = .93$) and 2-week test–retest ($r = .90$) reliability, and have correlated in predicted directions with criterion variables such as career decision making, vocational identity, and optimism (Juntunen & Wettersten, 2006). Total scores for the WHS were examined in the present study.

Prosocial work motivation. Four items developed by Grant (2008) to measure prosocial work motivation were utilized in the present study. After being prompted with the question, ''Why are you motivated to do your work?'' participants are instructed to rate the extent to which each of four statements (e.g., Because I care about benefiting others through my work) is true of them, using a 7-point scale ($1 = $ *disagree strongly*, $7 = $ *agree strongly*). Grant (2008) reported an internal consistency reliability coefficient of .90 for scores on the scale. Furthermore, scores were found to be distinct (using factor analytic procedures) from intrinsic motivation scores, were positively correlated with intrinsic motivation and job satisfaction scores, and were not significantly correlated with job performance or productivity (Grant, 2008).

Meaning in life. Steger, Frazier, Oishi, and Kaler's (2006) Meaning in Life Questionnaire (MLQ) is comprised of 10 items prompting participants to think about what in life they feel is important. Using a 7-point scale ($1 = $ *absolutely untrue*, $7 = $ *absolutely true*), MLQ items are used to compute scores for the presence of meaning (αs ranging from .81 to .86) and search for meaning (αs ranging from .84 to .92) with 1-month test–retest reliabilities of .70 and .73, respectively. MLQ scores also have been shown to correlate in predicted directions with criterion variables using both self-report and informant reports (Steger, Frazier, Oishi, & Kaler, 2006).

Results

Correlations among study variables across self-report and informant-report methods are reported in Table 4. Reliability coefficients also are reported in Table 4 and provide strong reliability evidence for scores on all study variables assessed using both methods.

Convergent and Discriminant Validity for CVQ-Presence and BCS-Presence Scores

Convergent correlations. Convergent validity is indicated when correlations of CVQ responses provided by self-report and informant-report are positive and substantial (heteromethod), and when CVQ scores correlate with responses on other measures of calling within self- and informant-reports (monomethod). Self-reports and informant-reports for CVQ-Presence scores correlated $r = .51$, providing evidence of convergent validity (heteromethod). The Self/Informant correlation for BCS-Presence was weaker but significant ($r = .27$), and the corresponding correlation for the calling paragraph was nonsignificant ($r = .15$). Similarly, CVQ-Presence and BCS-Presence scores correlated positively and significantly within self-reports and informant-reports (monomethod; $r = .69$ for self-report, $r = .42$ for informant-report). This same pattern was observed for correlations between the CVQ-Presence and the BCS-Presence using the heteromethod approach (r's of .46 and .32). Correlations between CVQ-Presence scores and the single-item calling paragraph scores were weaker; the four pairings of these variables ranged from $r = .11$ to $r = .38$, and two were nonsignificant. Correlations between BCS-Presence and the calling paragraph were similar (r's from .04 to .38), suggesting that the calling paragraph assessed a different construct than did CVQ-Presence and BCS-Presence scales.

Discriminant correlations. Discriminant validity is indicated when the correlations identified above exceed those observed between the CVQ and other constructs (heterotrait).

Evidence for discriminant validity was supported in that the magnitudes of correlations between CVQ-Presence and BCS-Presence scores (i.e., convergent correlations) generally were higher than correlations found between scores on either scale with scores on work hope, prosocial work motivation, meaning in life, and life satisfaction (i.e., discriminant correlations). Specifically, the monomethod convergent correlation among self-report scores on the CVQ-Presence with BCS-Presence of .69 exceeded all monomethod correlations of scores from these two calling measures with scores from other heterotrait measures (r's of .01 to .60).

The heteromethod convergent correlation between self- and informant-report scores on the CVQ Presence scale was .51. This exceeded all heteromethod discriminant correlations (r's of −.04 to .31). The heteromethod convergent correlation between self- and informant-report scores on the BCS Presence was .27. This exceeded some heteromethod discriminant correlations but was lower than correlations of self-report BCS-Presence scores with informant-report scores on the calling paragraph ($r = .28$), work hope scale ($r = .28$), and prosocial work attitudes scale ($r = .30$). The monomethod convergent correlation between informant-report BCS and CVQ scores was .42, which exceeded all discriminant correlations except the correlations of informant-report scores on the CVQ-Presence with prosocial work motivation ($r = .62$) and CVQ-Search ($r = .58$).

Convergent correlations for the calling paragraph with CVQ-Presence and BCS-Presence computed using monomethod (i.e., both self-report or both informant-report) scores were larger than all discriminant correlations except for its correlation with work hope ($r = .29$) for self-report scores. However, because three of the four heteromethod convergent correlations with the calling paragraph were small and nonsignificant, ranging from r's of .04 to .18, discriminant validity was not supported for these scores.

Table 4. Convergent and Discriminant Correlations Using Self- and Informant-Report Methods

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. s-CVQ Presence	(.88)																			
2. s-BCS Presence	.69**	(.79)																		
3. s-Calling paragraph	.27**	.24**	—																	
4. s-Work Hope	.35**	.34**	.29**	(.93)																
5. s-Prosocial	.54**	.25**	.17	.43**	(.98)															
6. s-MLQ Presence	.50**	.59**	.23*	.55**	.38**	(.86)														
7. s-Life Satisfaction	.08	.01	-.01	.31**	.11	.37**	(.83)													
8. s-CVQ Search	.60**	.29**	.03	.04	.45**	.16	.00	(.85)												
9. s-BCS Search	.14	.02	-.10	-.23*	.03	-.19*	-.07	.46**	(.82)											
10. s-MLQ Search	.13	.02	-.07	-.15	.18*	-.14	-.12	.50**	.46**	(.86)										
11. i-CVQ Presence	.51**	.46**	.11	.29**	.30**	.31**	-.04	.23**	.07	.06	(.87)									
12. i-BCS Presence	.32**	.27**	.04	-.02	.07	.17	.05	.14	.07	.07	.42**	(.82)								
13. i-Calling paragraph	.18	.28**	.15	.50**	.14	.36**	.27**	-.04	-.15	-.14	.38**	.38**	—							
14. i-Work Hope	.18	.28**	.15	.50**	.14	.36**	.27**	-.04	-.15	-.15	.38**	.38**	.23**	(.94)						
15. i-Prosocial	.31**	.30**	.14	.25**	.29**	.25**	.04	.15	-.07	.03	.62**	.24**	.34**	.49**	(.96)					
16. i-MLQ Presence	.11	.13	-.03	.17	.00	.23**	.36**	.03	-.11	-.08	.28**	.15	.12	.51**	.24**	(.89)				
17. i-Life Satisfaction	.09	.06	-.16	.14	-.01	.19*	.33**	.01	-.08	-.04	.27**	.23**	.12	.55**	.25**	.89**	(.91)			
18. i-CVQ Search	.26**	.16	-.01	.03	.30**	.11	-.11	.36**	.19**	.33**	.58**	.18*	.17*	.02	.48**	.00	.02	(.87)		
19. i-BCS Search	.07	.03	.01	-.18*	.05	-.01	-.07	.25**	.37*	.21	.08	.10	-.02	-.31**	-.03	-.26**	-.23**	.48**	(.85)	
20. i-MLQ Search	.03	.12	-.12	.10	-.03	.27**	.37**	.03	-.06	-.04	.22**	.19*	.12	.39**	.23**	.79**	.77**	.05	-.15	(.91)

Note. i = informant-report; s = self-report. () = Coefficient α values reported for each scale with 3 or more items, correlation coefficients reported for BCS scales, no value reported for the calling paragraph, as it is a single item scale. CVQ Presence = Calling and Vocation Questionnaire, presence total score; BCS Presence = Brief Calling scale, presence score; Calling paragraph = Wrzesniewski et al's (1997) single-item scale; Work Hope = Juntunen and Wettersten's (2006) Work Hope scale; Prosocial = Grant's (2008) Prosocial Work Orientation scale; MLQ Presence = Meaning in Life Questionnaire, Presence scale; CVQ Search = Calling and Vocation Questionnaire, Search total score; BCS Search = Brief Calling scale, Search score; MLQ Search = Meaning in Life Questionnaire, Search scale.

*p < .05. **p < .01.

Convergent and Discriminant Validity for CVQ-Search and BCS-Search Scores

Convergent correlations. Heteromethod convergent correlations between self-report and informant-report scores were significant and positive for both the CVQ-Search ($r = .36$) and the BCS-Search ($r = .37$). Monomethod convergent correlations between CVQ-Search and BCS-Search scores also were positive and significant within self-report ($r = .46$) and informant-report ($r = .48$) scores. Heteromethod convergent correlations of self-reports with informant-reports were smaller, but still significant, correlations (rs of .19 and .25). Nonsignificant correlations were found between CVQ-Search and BCS-Search with the calling paragraph in all cases except monomethod informant-reports, which yielded $r = .17$, again suggesting differences in the constructs measured by the calling paragraph with the CVQ and BCS search scales.

Discriminant correlations. Evidence for discriminant validity of CVQ-Search and BCS-Search scores was supported in that the monomethod convergent correlation of self-report scores was stronger than correlations of CVQ-Search scores with other variables, with the exception of an $r = .50$ correlation with MLQ-Search self-report scores. The same pattern was found for BCS-Search scores. Results were similar with monomethod informant-report scores, except for the correlations among informant-reports of the CVQ-Search with prosocial work motivation scores ($r = .50$) and CVQ-Presence scores ($r = .58$). Heteromethod convergent correlations among self-report CVQ-Search scores exceeded all self-report discriminant correlations. Heteromethod convergent correlations among informant-report CVQ-scores exceeded most informant-report discriminant correlations, except for correlations between informant-report CVQ-Search and both self-report CVQ-Presence ($r = .26$) and prosocial work motivation ($r = .30$) scores. For BCS-Search, heteromethod convergent correlations among both self-report and informant-report scores exceeded all discriminant correlations. Because convergent correlations for the calling paragraph with CVQ-Search and BCS-Search were very low, discriminant validity for its scores could not be established.

Summary and Discussion

Results from the multitrait, multimethod analysis generally supported convergent and discriminant validity for CVQ and BCS scores representing the presence of, and search for, a calling. For both instruments, convergent validity evidence was somewhat stronger for Presence scores compared to Search scores. Discriminant validity evidence was supported by the predicted patterns of relationships, with a few departures. The weight of this evidence, paired with the strong evidence of internal consistency reliability, supports the practice of summing the three presence subscales and three search subscales to form total scores for CVQ-Presence and CVQ-search, respectively. The strong relationships found between CVQ scores with prosocial motivation scores under some conditions are not surprising, given the conceptual similarity of the third dimension of Dik and Duffy's (2009) definition of calling with prosocial work motivation. The high correlations between presence and search dimensions for the CVQ, which were notably higher than those between the presence and search dimensions for the BCS, suggest that although participants may differentiate between seeking and experiencing when thinking about calling, generally they appear less likely to do so when evaluating the three dimensions of calling in a more specific and multidimensional manner. This supports the interpretation that calling is a process in which people experience, maintain, and seek their callings on a continuous and ongoing basis; that is, part of having a calling may mean continually seeking ways to maintain, enhance, or expand that calling. Finally, results suggested that BCS scores were supported at a level comparable to that associated with total scores on the CVQ, but that calling paragraph scores were not well-supported by validity evidence.

General Discussion

The primary purpose of the current research was to address a gap in the literature by developing and establishing psychometric support for a multidimensional measure of calling rooted in a clear definition of the construct. The resulting CVQ yields scores for both the presence of, and search for, a sense of calling, as well as three subscale scores for the presence dimension and three for the search dimension. Evidence from Study 1 supported the measurement model and revealed strong evidence for internal consistency reliability, while demonstrating a moderate level of stability from 1-month test–retest reliability coefficients for both subscale and total scores. Furthermore, the study provided initial evidence supporting criterion-related and construct validity of CVQ-Presence and CVQ-Search scores. Study 2 provided a more comprehensive and sophisticated means of evaluating construct validity for CVQ total scores; it did so for BCS scores as well, thus satisfying a secondary purpose of the current research. Strong evidence of internal consistency reliability was found for both instruments' scores, and results supported the construct validity of both CVQ and BCS scores, which performed comparably to each other and favorably to scores obtained using the single-item calling paragraph. To summarize, although additional research is needed, these studies establish the CVQ and affirm the BCS as psychometrically supported means of assessing calling within one's work or career.

Both the BCS and CVQ offer improvements over existing measures of calling. The BCS assesses presence and search dimensions with just 4 items, an important advantage for researchers interested in measuring these general dimensions of calling with maximum efficiency. The CVQ is less efficient in terms of number of items, but is the first instrument of which we are aware that provides a multidimensional approach to assessing calling in a manner tied to a clearly articulated definition of the construct. The CVQ also provides users with the flexibility to examine calling at the level of total scores for presence and search, as well as at the more specific subscale level. For example, users also could explore differences between individuals who resonate with all three components of a calling with those who find a transcendent summons irrelevant, but are nevertheless well-described with the second and third dimensions of the term. To use Dik and Duffy's (2009) terminology, such a study would compare those with a calling to those with vocations, but not callings. The CVQ also can be used, potentially, across a wide range of populations. For example, because it refers to one's work or career generally, it can be used with participants representing different career paths, and because it refers to a transcendent summons but leaves open the source of the summons, it can be used with participants representing a range of religious, spiritual, and cultural worldview beliefs.

Although presence and search subscale scores across the CVQ and BCS were highly correlated, the fact that CVQ presence and search scores were strongly positively correlated, while BCS presence and search scores were weakly correlated, suggests that different processes may be at work in how participants evaluate calling items that require general appraisals compared to those that assess more specific facets of calling. Similar to speculation for why facet measures of job satisfaction often do not equal ratings of overall job satisfaction (Fritzsche & Parrish, 2005), it may be that different people who identify with the concept of calling may emphasize different dimensions of the domain. As noted earlier, the correlation of CVQ Presence and Search scores may suggest that calling is an ongoing process rather than a one-time event. For example, instead of someone seeking a calling but not yet experiencing one, then ceasing the search upon discovering her or his calling, perhaps people oriented to the concept of calling are continually and dynamically trying to evaluate their careers and approach their work in a way that maintains or increases the experiential properties that reflect a calling. Such possibilities have not yet been addressed in research on calling, and are fruitful areas to explore.

Future research also could also investigate the six subscales of the CVQ. Although the cross-validated measurement model in Study 1 supported the presence of six factors, space constraints precluded a more thorough evaluation of the subscales in this article. Relevant questions that can be explored include: To what extent does each subscale add incremental validity when assessing overall

presence of, or search for, a calling? To what extent does the transcendent summons dimension behave differently for those with sacred versus secular conceptualizations of calling (see Steger et al., 2010)? Are purposeful work and prosocial orientations differentially active for individuals within individualist versus collectivist cultures? Such questions could not be readily addressed prior to the development of the CVQ.

The transcendent summons, purposeful work, and prosocial orientation domains of calling have served as a means of organizing recommendations for career counseling practice designed to address and integrate clients' perceptions of calling (Dik et al., 2009). Although counselors can use informal and qualitative means of assessing these calling components, the CVQ can corroborate these efforts, and identify the relative salience of particular dimensions of calling to be addressed in counseling. The CVQ or BCS may also serve as an entry point for discussing calling with clients, or as a screening tool to identify the relevance of calling among clients at the outset of counseling. See Dik, Duffy, and Eldridge (2009) for a more thorough discussion of practice recommendations relevant to the three calling domains. Although organizational applications need to be further developed, managers and human resource professionals may also find the CVQ helpful as a screening tool or as part of a battery of assessments to measure the psychological climate within a work unit or organization.

Limitations

Other directions for future research stem from several limitations of the current research, which should be weighed when using the CVQ or BCS. First, the samples in the current research were disproportionately White and female. Although a recent study found no differences in the CVQ measurement model across gender (Eldridge, 2010), further research is needed that examines the generalizability of the instrument across different groups. Second, the sample with whom the CVQ item pool was refined and tested consisted of college students. Although calling is highly developmentally relevant for college students (e.g., Hunter et al., 2010), it cannot be assumed that nonstudent adults approach the construct in the same way; this is a matter for future research to address. Third, participants in the sample were drawn from two regions in the United States. Future research should explore possible regional, national, and cultural differences that exist in how calling is conceptualized and in the degree to which it is considered a relevant construct for participants. Fourth, although CVQ and BCS scores were found related to a wide array of criterion variables in the current research, the variables and scales included were by no means exhaustive. Future research should compare the CVQ and BCS to additional measures of calling, and more comprehensively investigate the nomological net of calling. Finally, it bears reminding that causal inferences cannot be made from correlational data, such as that presented in the current research. Research is needed that examines how a sense of calling may develop over time, and that investigates the nature and extent of causal influences that calling may have on work-related and well-being outcomes.

Appendix A
Calling and Vocation Questionnaire (CVQ)

Instructions: Please indicate the degree to which you believe the following statements describe you, using the following scale. Please respond with your career as a whole in mind. For example, if you are currently working part time in a job that you don't consider part of your career, focus on your career as a whole and not your current job. Try not to respond merely as you think you "should" respond; rather, try to be as accurate and as objective as possible in evaluating yourself. If any of the questions simply do not seem relevant to you, "1" may be the most appropriate answer.

> 1 = Not at all true of me
> 2 = Somewhat true of me
> 3 = Mostly true of me
> 4 = Absolutely true of me

1. I believe that I have been called to my current line of work.
2. I'm searching for my calling in my career.
3. My work helps me live out my life's purpose.
4. I am looking for work that will help me live out my life's purpose.
5. I am trying to find a career that ultimately makes the world a better place.
6. I intend to construct a career that will give my life meaning.
7. I want to find a job that meets some of society's needs.
8. I do not believe that a force beyond myself has helped guide me to my career.
9. The most important aspect of my career is its role in helping to meet the needs of others.
10. I am trying to build a career that benefits society.
11. I was drawn by something beyond myself to pursue my current line of work.
12. Making a difference for others is the primary motivation in my career.
13. I yearn for a sense of calling in my career.
14. Eventually, I hope my career will align with my purpose in life.
15. I see my career as a path to purpose in life.
16. I am looking for a job where my career clearly benefits others.
17. My work contributes to the common good.
18. I am trying to figure out what my calling is in the context of my career.
19. I'm trying to identify the area of work I was meant to pursue.
20. My career is an important part of my life's meaning.
21. I want to pursue a career that is a good fit with the reason for my existence.
22. I am always trying to evaluate how beneficial my work is to others.
23. I am pursuing my current line of work because I believe I have been called to do so.
24. I try to live out my life purpose when I am at work.

Scoring instructions (items listed should be summed):

   CVQ-Presence-Transcendent Summons = 1, 8-reverse coded, 11, 23
   CVQ-Search-Transcendent Summons = 2, 13, 18, 19
   CVQ-Presence-Purposeful Work = 3, 15, 20, 24
   CVQ-Search-Purposeful Work = 4, 6, 14, 21
   CVQ-Presence-Prosocial Orientation = 9, 12, 17, 22
   CVQ-Search-Prosocial Orientation = 5, 7, 10, 16
   CVQ-Presence total = 1, 3, 8-reverse coded, 9, 11, 12, 15, 17, 20, 22, 23, 24
   CVQ-Search total = 2, 4, 5, 6, 7, 10, 13, 14, 16, 18, 19, 21

Appendix B
Brief Calling Scale (BCS)

Broadly speaking, a "calling" refers to a person's belief that she or he is called upon (by the needs of society, by a person's own inner potential, by God, by a Higher Power, etc.) to do a

particular kind of work. The following questions assess the degree to which you see this concept as relevant to your own life and career. Please respond honestly, not according to what is socially desirable or what you feel you "ought" to think. Please indicate the extent to which each of the following statements <u>currently</u> describe you, using the following scale.

> 1 = Not at all true of me
> 2 = Mildly true of me
> 3 = Moderately true of me
> 4 = Mostly true of me
> 5 = Totally true of me

1. I have a calling to a particular kind of work.
2. I have a good understanding of my calling as it applies to my career.
3. I am trying to figure out my calling in my career.
4. I am searching for my calling as it applies to my career.

Scoring instructions (items listed should be summed):

BCS-Presence = 1, 2
BCS-Search = 3,4.

Author's Note

Brandy M. Eldridge is currently with the University of New Mexico, Albuquerque, NM, USA.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

References

Amabile, T. M., Hill, K. G., Hennessey, B. A., & Tighe, E. M. (1994). The work preference inventory: Assessing intrinsic and extrinsic motivational orientations. *Journal of Personality and Social Psychology*, *66*, 950–967. doi:10.1037/0022-3514.66.5.950

Arbuckle, J. L. (2005). *AMOS 6.0 [Software]*. Chicago, IL: Small Waters Corp.

Bellah, R. N., Madsen, R., Sullivan, W. M., Swidler, A., & Tipton, S. M. (1986). *Habits of the heart: Individualism and commitment in American life.* New York, NY: Harper & Row.

Betz, N. E., Hammond, M. S., & Multon, K. D. (2005). Reliability and validity of five-level response continua for the career decision self-efficacy scale. *Journal of Career Assessment*, *13*, 131–149. doi:10.1177/1069072704273123

Bunderson, J. S., & Thompson, J. A. (2009). The call of the wild: Zookeepers, callings, and the dual edges of deeply meaningful work. *Administrative Science Quarterly*, *54*, 32–57.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*, 309–319.

Dalton, J. C. (2001). Career and calling: Finding a place for spirit in work and community. *New Directions for Student Services*, *95*, 17–25.

Davidson, J. C., & Caddell, D. P. (1994). Religion and the meaning of work. *Journal for the Scientific Study of Religion*, *33*, 135–147.

Dawis, R. V. (2000). Scale construction and psychometric considerations. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 65–94). San Diego, CA: Academic Press.

Diener, E., Emmons, R. A., Larson, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*, 71–75.

Dik, B. J., & Duffy, R. D. (2009). Calling and vocation at work: Definitions and prospects for research and practice. *The Counseling Psychologist*, *37*, 424–450. doi:10.1177/0011000008316430

Dik, B. J., Duffy, R. D., & Eldridge, B. M. (2009). Calling and vocation in career counseling: Recommendations for promoting meaningful work. *Professional Psychology: Research and Practice*, *40*, 625–632. doi:10.1037/a0015547

Dik, B. J., Duffy, R. D., & Steger, M. F. (2012). Enhancing social justice by promoting prosocial values in career development interventions. *Counseling and Values, 57,* 31–37. doi: 10.1002/j.2161-007X.2012.00005.x

Dik, B. J., Sargent, A. M., & Steger, M. F. (2008). Career development strivings: Assessing goals and motivation in career decision-making and planning. *Journal of Career Development*, *35*, 23–41. doi:10.1177/0894845308317934

Dobrow, S. R., & Tosti-Kharas, J. (2011). Calling: The development of a scale measure. *Personnel Psychology, 64*, 1001–1049.

Dreher, D. E., Holloway, K. A., & Schoenfelder, E. (2007). The vocation identity questionnaire: Measuring the sense of calling. *Research in the Social Scientific Study of Religion*, *18*, 99–120.

Duffy, R. D., Allan, B. A., & Dik, B. J. (2011). The presence of a calling and academic satisfaction: Examining potential mediators. *Journal of Vocational Behavior*, *79*, 74–80. doi:10.1016/j.jvb.2010.11.001

Duffy, R. D., Dik, B. J., & Steger, M. S. (2011). Calling and work related outcomes: Career commitment as a mediator. *Journal of Vocational Behavior*, *78*, 210–218. doi:10.1016/j.jvb.2010.09.013

Duffy, R. D., & Sedlacek, W. E. (2007). The presence of and search for a calling: Connections to career development. *Journal of Vocational Behavior*, *70*, 590–601. doi:10.1016/j.jvb.2007.03.007

Duffy, R. D., & Sedlacek, W. E. (2010). The salience of a career calling among college students: Exploring group differences and links to religiousness, life meaning, and life satisfaction. *The Career Development Quarterly*, *59*, 27–40.

Elangovan, A. R., Pinder, G. C., & McLean, M. (2010). Callings and organizational behavior. *Journal of Vocational Behavior*, *76*, 428–440. doi:10.1016/j.jvb.2009.10.009

Eldridge, B. M. (2010). *Structure of calling and vocation across gender and age cohort (Unpublished doctoral dissertation)*. Colorado State University, Fort Collins, CO.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, *7*, 286–299.

Fritzsche, B., & Parrish, T. (2005). Theories and research on job satisfaction. In S. Brown & R. Lent (Eds), *Career development and counseling* (pp. 180–202). New York, NY: Wiley.

Grant, A. M. (2008). Does intrinsic motivation fuel the prosocial fire? Motivational synergy in predicting persistence, performance, and productivity. *Journal of Applied Psychology*, *93*, 48–58. doi:10.1037/0021-9010.93.1.48

Hardy, L. (1990). *The fabric of this world: Inquiries into calling, career choice, and the design of human work*. Grand Rapids, MI: Eerdmans.

Harvey, V. S. (1997). Improving readability of psychological reports. *Professional Psychology: Research and Practice*, *28*, 271–274.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, *7*, 191–205. doi:10.1177/1094428104263675

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.

Hunter, I., Dik, B. J., & Banning, J. (2010). College student's perception of calling in work and life: A qualitative investigation. *Journal of Vocational Behavior*, *76*, 178–186. doi:10.1016/j.jvb.2009.10.008

Juntunen, C. L., & Wettersten, K. B. (2006). Work hope: Development and initial validation of a measure. *Journal of Counseling Psychology*, *53*, 84–106. doi:10.1037/0022-0167.53.1.94

Kahn, J. H. (2006). Factor analysis in counseling psychology research, training, and practice: Principles, advances, and applications. *The Counseling Psychologist*, *34*, 684–718. doi:10.1177/0011000006286347

Pavot, W., & Diener, E. (1993). Review of the satisfaction with life scale. *Psychological Assessment*, *5*, 164–172. doi:10.1007/978-90-481-2354-4_5

Peterson, C., Park, N., Hall, N., & Seligman, M. E. P. (2009). Zest and work. *Journal of Organizational Behavior*, *30*, 161–172. doi:10.1002/job.584

Serow, R. C. (1994). Called to teach: A study of highly motivated preservice teachers. *Journal of Research and Development in Education*, *27*, 65–72.

Steger, M. F. (2006). An illustration of issues in factor extraction and identification of dimensionality in psychological assessment data. *Journal of Personality Assessment*, *86*, 263–272. doi:10.1207/s15327752jpa8603_03

Steger, M. F., & Dik, B. J. (2009). If one is searching for meaning in life, does meaning in work help? *Applied Psychology: Health and Well-Being*, *1*, 303–320.

Steger, M. F., Frazier, P., Oishi, S., & Kaler, M. (2006). The meaning in life questionnaire: Assessing the presence of and search for meaning in life. *Journal of Counseling Psychology*, *53*, 80–93. doi:10.1037/0022-0167.53.1.80

Steger, M. F., Pickering, N., Shin, J. Y., & Dik, B. J. (2010). Calling in work: Secular or sacred? *Journal of Career Assessment*, *18*, 82–96. doi:10.1177/1069072709350905

Treadgold, R. (1999). Transcendent vocations: Their relationship to stress, depression, and clarity of self-concept. *Journal of Humanistic Psychology*, *39*, 81–105.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070.

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, *34*, 806–838. doi:10.1177/0011000006288127

Wrzesniewski, A. (2010). Callings. In K. Cameron & G. Spreitzer (Eds.), *Handbook of positive organizational scholarship.* Oxford, UK: Oxford University Press.

Wrzesniewski, A., McCauley, C. R., Rozin, P., & Schwartz, B. (1997). Jobs, careers, and callings: People's relations to their work. *Journal of Research in Personality*, *31*, 21–33. doi:10.1006/jrpe.1997.2162