Journal of Sport and Exercise Psychology, 2008, 30, 561-587

Long and Short Measures of Flow: The Construct Validity of the FSS-2, DFS-2, and New Brief Counterparts

Susan A. Jackson,[1] Andrew J. Martin,[2] and Robert C. Eklund,[3]

[1]University of Queensland, [2]University of Sydney, and [3]Florida State University

Long and short flow scales are examined from dispositional (*n* = 652 long; *n* = 692 short) and state (*n* = 499 long; *n* = 865 short) perspectives. The long flow scales constitute a 36-item multidimensional assessment of flow and have previously demonstrated good psychometric properties. The short flow scales constitute new abbreviated versions of the long forms, contain 9 items, and provide a brief measure of flow from a dimensional perspective. In the current study, long and short flow scales are assessed across a large and diverse physical activity sample. With few exceptions, these flow measures demonstrated acceptable model fit, reliability, and distributions; associations with key correlates in parallel and hypothesized ways; and invariance in factor loadings. Together, the scales provide options for assessing flow in different contexts and when different goals or constraints are operating. Researchers wanting to capture an aggregate of the multidimensional framework might find the short scales a pragmatic alternative when constraints prohibit use of the full-length versions.

Keywords: flow, measurement, short scales, construct validity, positive psychology

Flow is an experience that both inspires and eludes. People, across diverse domains, strive to experience flow in their activities. At the same time, flow often eludes the seeker, presenting itself on relatively rare occasions. Flow has been identified as a key construct in the burgeoning field of positive psychology (e.g., Nakamura & Csikszentmihalyi, 2002). It is characterized by absorption in what one is doing, and as a result of this centeredness of mind, one's subjective experience is optimized. Indeed, flow is one of the key capacities identified by positive psychology (Seligman & Csikszentmihalyi, 2000)—a capacity that can be developed through building appropriate competencies. Csikszentmihalyi (e.g., 1990) described in detail the factors that lead to and form the structure of the flow

Jackson is with the School of Human Movement Studies, University of Queensland, Brisbane, Australia; Martin is with the Faculty of Education and Social Work, University of Sydney, Sydney, Australia; and Eklund is with the Department of Educational Psychology and Learning Systems, Florida State University, Tallahassee, FL. Copyright to the long and short flow scales is held by Jackson. Please direct any enquiries regarding use of the scales to the authors.

experience. From this ground-breaking work, the key elements of flow have been identified. Using these elements as a foundation, the present research continues the examination of measurement tools for researchers interested in assessing dimensions of flow.

Previous Flow Assessment

One of the greatest challenges in flow research—or any research involving subjective experiences—is finding ways to accurately and reliably assess the experience itself. Qualitative methods have provided rich accounts of flow, and this was the first approach, via semistructured interviews, to understanding this phenomenon (Csikszentmihalyi, 1975). Subsequent in-depth qualitative investigations of flow in sport have made apparent the relevance of the construct to athletes (e.g., Jackson, 1992, 1996). Notwithstanding the rich information provided by qualitative methods, quantitative approaches are also widely used in flow research. Most notably, Csikszentmihalyi and colleagues (e.g., Csikszentmihalyi & Larson, 1987) developed an approach designed to capture experience as it was occurring: the Experience Sampling Method (ESM). The ESM is typically completed over a period of 1 week, to provide an experiential picture of the respondent's daily life. While the ESM offers many advantages and is the tool of choice when the goal is to examine experience in situ, there is also potential in more standard forms of self-report instrumentation to assess components of the flow experience. For example, when it is not pragmatic to interrupt people during their activity, or to sample experience repeatedly across the course of a week, standard questionnaire techniques may be preferable.

Accordingly, Jackson and colleagues (Jackson & Eklund, 2002; Jackson & Marsh, 1996) developed the Flow State Scale (FSS) and a corresponding dispositional version of the instrument, the Dispositional Flow Scale (DFS). The DFS is a dispositional assessment of the frequency with which people experience flow in a target activity. The FSS was designed to be completed after a specific event, to assess the state, or situation-specific experience of flow. These measures were developed and validated in physical activity settings. However, they have been used in other performance-related domains as well (e.g., Jackson & Eklund, 2004; Martin & Cutler, 2002).

In both the DFS and FSS, nine key flow characteristics, or dimensions, are assessed, based on the nine-dimensional conceptual flow model (Csikszentmihalyi, 1990; Jackson & Csikszentmihalyi, 1999): challenge-skill balance, action-awareness merging, clear goals, unambiguous feedback, concentration on the task at hand, sense of control, loss of self-consciousness, time transformation, and autotelic experience. These instruments underwent stringent psychometric examination (Jackson & Marsh, 1996; Marsh & Jackson, 1999). Although both first-order (nine-factor) models and a higher order model involving a global flow factor were found to have acceptable reliability and factor structure, there were areas where improvements could be made. Item modifications were made to the DFS and FSS to improve the measurement of some of the flow dimensions (Jackson & Eklund, 2002). Confirmatory factor analyses supported the construct validity of the new scales (i.e., DFS-2, FSS-2) and acceptable internal consistency (subscale alphas ranging from .78 to .92) in two independent samples. In sum, these flow scales provide a valid and reliable means of assessing flow, and have been used to do so in a range of activities and settings.

Jackson and Eklund's (2002) validation studies showed support for the nine-factor first-order flow model. There was also satisfactory fit for the higher order model, although this higher order fit was attenuated by the relatively weaker loadings from two of the flow dimensions. The overall fit is marginally weaker than the nine-factor first-order model—but this is to be expected when comparing a first-order model with freely estimated correlations to a higher order model that is constrained to have all first-order factor intercorrelations explained by the one factor. While the generally recommended procedure for use of the DFS-2 and FSS-2 is to use factor level scores rather than a single global flow score, it is also recognized that there are some situations where the research question means that the use of one rather than nine factor scores may be preferable. Jackson and colleagues have addressed this area in earlier construct validity studies (see Jackson & Eklund, 2002; Jackson & Marsh, 1996; Marsh & Jackson, 1999).

A Brief Version of the Flow Scales

Psychometric research emphasizes the importance of measurement based on multiple items for a given scale (e.g., Marsh, Martin, & Hau, 2006) to enable more reliable assessment of target constructs and a capacity to model systematic and random error (Williams, Ford, & Nguyen, 2002). Nonetheless, administrative constraints often dictate a need for more succinct forms. For example, in multitrait multimethod research (Marsh, Martin, & Hau, 2006), "observer" or "other" ratings may be used (e.g., coach rating athlete) and, because these observers often must rate many individuals, a short form is desirable. Alternatively, in large-scale projects involving many measures, short forms may be preferred to alleviate participant burden. Where flow is not a central construct, a shorter form may provide an opportunity to assess the experience without imposing too many restrictions on measurement of more central constructs. Similarly, when there are time constraints making long instrumentation difficult to employ, short forms may be preferable. For reasons such as these, the short (nine-item) flow measure was developed—a summary scale of the long multidimensional one—broadly reflecting the higher order factor structure. The present study concurrently assesses the validity and reliability of both the full-length flow scales and these new brief forms. In this manuscript, we refer to the original multidimensional flow scales (i.e., the DFS-2 and FSS-2) as the *long scales*, whereas the abbreviated nine-item versions (broadly in line with the higher order structure of the long scales) examined in this manuscript are referred to as the *short scales.*

The idea for a short flow scale was driven by the perceived need, as described above, to develop a succinct measure to use when there might exist research or practical constraints preventing the use of the full-length measure of flow (i.e., DFS-2, FSS-2). Given the extant support for the higher order model in the long scales, it was deemed worthwhile to examine whether a shorter version of the flow scales could provide a valid and reliable measure of this global flow construct. To develop the short scale, the first author selected a set of nine items from the long flow scales—one for each of the nine flow dimensions. It was considered critical to include an item from each of the nine flow factors in the short scale, to retain the conceptual model of flow (Csikszentmihalyi, 1990) in this abbreviated measure. Items were selected on two criteria. First, on empirical grounds, items were identified that appeared to best measure the intended construct. Items were selected based on the size of their standardized factor loadings in previous CFA (Jackson & Eklund, 2002; Jackson & Marsh, 1996) from each of the nine flow factors comprising the long flow scales (Marsh, Ellis, Parada, Richards, & Heubeck, 2005). Second, where items representing the same dimension had equal or very close factor loadings, items that were deemed to have better face validity and provided a good overall representation of that dimension were chosen.

Martin, Tipler, Marsh, Richards, and Williams (2006) trialed the dispositional short flow scale in a sample of high school students where relationships with measures of motivation in physical activity as well as self-reported physical activity levels were assessed. The short DFS-2 yielded acceptable reliability (alpha = .82). Predicted relationships with physical activity motivation were found, including associations with adaptive cognitions and physical activity behaviors. This study provided good preliminary support for the short flow scale, but was restricted to the dispositional version of the instrument and to one sample of high school students. Martin (2008b, in press) also conducted research with the short flow scale, using short flow as one of a set of dependent measures in two domain specificity studies of motivation and engagement in music and sport. Further research is needed to more specifically examine the psychometric properties of the short flow scales, and to do so in the context of physical activity, from which the short scales' predecessors (i.e., the DFS-2, FSS-2) originated. This study provides a comprehensive examination of both dispositional and state versions of long and short flow scales in a large, diverse physical activity sample, using confirmatory factor analyses.

A Construct Validation Approach to Assessing Flow

Psychometric researchers have increasingly emphasized the need to both develop and evaluate instruments within a construct validation framework (e.g., Marsh, 1997). Studies that adopt a construct validation approach can be classified as within-network or between-network studies. Within-network studies explore the internal structure of a construct. Beginning with a logical analysis of internal consistency of the construct definition, measurement instruments, and generation of predictions, within-network studies typically employ empirical techniques such as confirmatory factor analysis (CFA) and reliability analysis. Between-network studies attempt to establish a logical, theoretically consistent pattern of relations between constructs. In reality, few construct validation studies adopt this dual (within- and between-network) approach and hence they provide relatively limited input into understanding the constructs and measures under focus. The present study uses both construct approaches. Firstly, it conducts a within-network study using confirmatory factor analysis to test the psychometric properties of the flow constructs in a large and diverse physical activity sample. Secondly, it conducts a between-network study by (a) examining the empirical links between the flow constructs and a set of between-network constructs (e.g., intrinsic motivation, self-concept, well-being) and (b) examining the invariance of factor structure across scales.

Between-Network Correlates for Flow

The between-network construct validation of the long and short versions of the flow scales in this study involved some constructs employed in previous validation studies, as well as some new constructs. This allowed us to extend the flow scale research in three ways: (1) by examining the generalizability of constructs previously demonstrated to be related to flow in a new, large sample; (2) by examining these relationships simultaneously in relation to the existing long flow scale as well as the new short flow scale; and (3) by examining some new hypothesized between-network correlates. The following between-network constructs provided a theoretically relevant basis for examining the validity of flow.

The first between-network construct examined was intrinsic motivation. Conceptually, flow and intrinsic motivation have been closely linked, with Csikszentmihalyi (1988) indicating his early interest in flow was sparked by wanting to understand what it felt like to be in an intrinsically motivating situation. In the current study, the measurement of intrinsic motivation was drawn from Deci and Ryan's (1985) self-determination theory (SDT), in which intrinsically motivated behavior is characterized as being engaged in for its own sake, rather than for any external reward. Previous research has found positive relationships between flow and intrinsic motivation in the physical activity domain (Jackson, Kimiecik, Ford, & Marsh, 1998; Martin et al., 2006), and these relationships were expected to be replicated in the current study.

Self-concept is widely regarded as a mediating variable across many domains. Here self-concept is encompassed by *sport self-concept, physical self-concept,* and *general self-esteem* (Marsh, Richards, Johnson, Roche, & Tremayne, 1994). Previous research, assessing dimensions of self-concept (Jackson, Thomas, Marsh, & Smethurst, 2001) different from those of the current study, found strong relationships between facets of multidimensional self-concept and flow. A related concept, *perceived competence* in one's targeted activity, is also included, as previous research by Jackson and colleagues (e.g., Jackson et al., 1998) has shown this construct to be strongly associated with flow. Having a high perception of one's abilities is a critical component to the potential to experience flow (Jackson & Csikszentmihalyi, 1999), and thus positive associations were expected between measures of self-concept/competence and flow.

Finally, *psychological well-being* was included as a between-network correlate. Four factors comprise psychological well-being in the present investigation: *anxiety, positive well-being, psychological distress,* and *fatigue*. *Anxiety/psychological distress* are proposed as opposing experiential states to flow in Csikszentmihalyi's (1988, 1990) conceptualization of the construct. Previous research has demonstrated negative associations between anxiety and flow (e.g., Jackson et al., 1998), and in the current study it was expected that anxiety and a new between-network construct of psychological distress would correlate negatively with the flow measures. An opposing construct to these negative experiential states, and also a new between-network construct, *positive well-being* as a measure of positive subjective experience, was expected to correlate positively with flow. *Fatigue*—part of the measure of subjective experience used in this study (McAuley & Courneya, 1994) and providing a new between-network construct, but tangential theoretically to flow—was not expected to correlate with flow.

Aims of the Present Investigation

The fundamental aim in the present investigation is to concurrently empirically assess an expanded flow measurement framework that encompasses (a) the standard multidimensional aspects of flow, as measured by the long flow scales, and (b) a unidimensional approach to flow, as measured by the short flow scales. We do so using a construct validity approach comprising extensive within-network validity (i.e., descriptive statistics and model fit) and between-network validity (i.e., relationships with key correlates and invariance tests).

From a within-network perspective, we hypothesize that both long and short flow measures will be reliable, approximately normally distributed, and evince acceptable model fit. By examining the long and short scales simultaneously in the one study, flow measures with robust psychometric support (the long DFS-2 and FSS-2) are able to be used to benchmark the psychometric properties of the new, short measures. From a between-network perspective, we hypothesize that long and short flow measures will relate to a set of key correlates and demonstrate invariance of factor structure across samples and scales. By including the same set of between-network correlates for the long and short flow scales, we are able to examine the relative between-network validity of the long and short versions. Finally, the inclusion of invariance testing extends previous flow scale construct validity, as few previous physical-activity studies have included invariance testing of the flow scales. In this study, we focus on invariance tests between the long and short scales, as well as between the dispositional and state versions of the scales. To the best of our knowledge, this is the first time such a comprehensive approach to examining construct validity of the long/short, and dispositional/state flow measures via invariance testing has been undertaken.

In summary, our overall aim is to show that these flow measures are psychometrically sound and appropriate for inclusion in research focused on understanding individuals' immersion and engagement with their activity. We also seek to offer some guidance to assist decision making about the most appropriate versions of the flow scales to use for various research purposes.

Method

Sample and Procedure

The total sample comprised 1,653 participants from a major capital city in Australia who were involved in some form of physical activity. Participants were asked to reflect on their experiences during participation in their self-selected activities. The sample consisted of 37% males and 62% females. The mean age of participants was 26 years (*SD* = 10.55), with 25% in late adolescence (less than 20 years of age), 48% in early adulthood (20–30 years), 19% in mid-adulthood (30–40 years 12%; 40–50 years 7%), and 6% in later adulthood (50 years plus).

The sample comprised a large number of groups and activity types (as summarized below). Accordingly, the survey type varied as a function of group. There were 652 long dispositional flow responses and 499 long state flow responses. For the short flow scales, there were 692 dispositional responses and 865 state responses. Some participants completed more than one survey type, with 580 responses to both the long and short dispositional scales and 475 responses to both the long and short state scales.

Just under half the sample (45%) indicated that their participation in their target activity was noncompetitive, while 55% indicated they competed at either club/school, state, national, or international level. There were 58 different activities listed by the participants. The most frequently represented target activities were netball (15%), yoga (13%), aerobics (8%), basketball (6%), hockey (6%), rugby (5%), dancing (5%), and running (5%). Average length of involvement in main activity was 6.6 years, with a range of 1 to 41 years. Over half (57%) of the sample participated in their activity two or more times per week, with 35% taking part once a week and 8% engaged in their activity less than once per week. Activities had to be classifiable as physical activity to be included in this sample. Standard informed consent procedures were followed in collection of the data, and institutional ethics approval for the research was obtained.

Materials

Long (36-Item) Flow Scales. The long flow scales are better known as the Dispositional Flow Scale 2 (DFS-2) and Flow State Scale 2 (FSS-2), developed by Jackson and Eklund (2002, 2004a), based on earlier instrumentation developed by Jackson and Marsh (1996). These 36-item scales are designed to assess flow experiences at dispositional (general experience in target activity) and state (experience in specific event in target activity) levels. These scales were theoretically grounded in Csikszentmihalyi's (1990) nine-dimensional conceptualization of flow, and this nine-factor structure has been supported through confirmatory factor analyses (Jackson & Eklund, 2002; Jackson & Marsh, 1996). A higher order model, representing a global flow construct, has also received reasonable confirmatory factor analytic support.

Both the DFS-2 and FSS-2 employ four items for each of the nine flow dimensions, with example items representing each dimension as follows: *Challenge-Skill Balance:* "I feel I am competent enough to meet the high demands of the situation"; *Action-Awareness Merging:* "I do things spontaneously and automatically without having to think"; *Clear Goals*: "I have a strong sense of what I want to do"; *Unambiguous Feedback:* "I have a good idea while I am performing about how well I am doing"; *Concentration on Task at Hand:* "I am completely focused on the task at hand"; *Sense of Control:* "I have a feeling of total control over what I am doing"; *Transformation of Time:* "The way time passes seems to be different from normal"; and *Autotelic Experience:* "The experience is extremely rewarding." The DFS-2 asks respondents to rate the frequency with which they experience the flow characteristics within a specified activity in general. The rating scale for the DFS-2 is a 5-point Likert scale, ranging from 1 (*never*) to 5 (*always*). Coefficient alpha estimates of reliability for the DFS-2 ranged from .78 to .90 in the Jackson and Eklund (2002) validation paper and in the current study are listed in Table 1, ranging from an acceptable .80 to .89. The FSS-2 asks respondents to rate the extent to which they experienced the flow characteristics on a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Reliability estimates for the FSS-2 in the Jackson and Eklund study ranged from .80 to .92 and in the current study are given in Table 1, ranging from an acceptable .76 to .90.

Short (9-Item) Flow Scales. The short dispositional and state flow scales each contain nine items, with one item from the four-item measures of each of the nine flow dimensions. The items selected for the short scales are described above in the example items for the long scales. The short scales employ the same response format as the corresponding 36-item scales described above. Preliminary analyses were conducted on the short scales using the nine items that formed part of the long 36-item set (embedded) from Jackson and Eklund's (2002) validation study. Coefficient alpha estimates of reliability from the item-identification and cross-validation data sets of Jackson and Eklund (2002) for the short DFS-2 were .81 and .74, and for the short FSS- 2 were .77 and .78, respectively. Goodness-of-fit indices based on these preliminary analyses of the short scales showed acceptable fit. The CFAs for the dispositional short flow for the item-identification sample (embedded $\chi^2 = 66.89$, $df = 27$; CFI = .99; NNFI = .98; RMSEA = .05; SRMR = .03) and the cross-validation sample (embedded $\chi^2 = 127.87$, $df = 27$; CFI = .96; NNFI = .95; RMSEA = .07; SRMR = .04) demonstrated a good fit of the hypothesized models to the data. Similarly, the CFAs for the state short flow for the item-identification sample (embedded $\chi^2 = 74.13$, $df = 27$; CFI = .97; NNFI = .97; RMSEA = .05; SRMR = .03) and the cross-validation sample (embedded $\chi^2 = 124.30$, $df = 27$; CFI = .97; NNFI = .96; RMSEA = .06; SRMR = .04) demonstrated a good fit of the hypothesized models to the data. Thus, the nine items selected for the short flow scales were assessed as providing a suitable short measure of flow.

Between-Network Correlates. Consistent with the between-network approach specified earlier, items were also administered that explored participant motivation (*intrinsic motivation)*, self-concept (*sport self-concept, physical self-concept, general self-esteem, perceived competence*), and psychological well-being (*positive well-being, anxiety, psychological distress, fatigue*). *Dispositional intrinsic motivation* (e.g., "For the excitement I feel when I am really involved in an activity"; Cronbach's $\alpha = .94$) from the Sport Motivation Scale (Pelletier et al., 1995) comprised 12 items that were rated on a scale of 1 (*does not correspond at all*) to 7 (*corresponds exactly*). The 12 items were originally designed to assess three subaspects of intrinsic motivation. In this study, they were combined into one intrinsic motivation measure—a practice that has been employed by the scale authors (i.e., Pelletier, Fortier, Vallerand, & Briere, 2001) and others (e.g., Amiot, Gaudreau, & Blanchard, 2004). A similar 7-point rating scale was used to assess *state intrinsic motivation*, with the Situational Motivation Scale (SIMS; Guay, Vallerand, & Blanchard, 2000), which asks participants to indicate why they are currently engaged in their activity (e.g., "Because I think that this activity is interesting"; Cronbach's $\alpha = .75$; four items). For self-concept, the *sport* (e.g., "I am good at most sports"; Cronbach's $\alpha = .95$; six items), *physical* (e.g., "Physically, I am happy with myself"; Cronbach's $\alpha = .96$; six items), and *general esteem* (e.g.,

Table 1 Long (36-Item) and Short (9-Item) Flow Descriptive and CFA Statistics

	Mean/5	*SD*	Skewness	Kurtosis	Reliability	CFA Load Range (mean)
DISPOSITION 36 item						
Challenge skill balance	3.69	.59	–.43	.56	.81	.50–.83 (.72)
Action awareness	3.74	.65	–.49	.76	.87	.76–.83 (.80)
Clear goals	3.97	.61	–.36	.63	.80	.63–.79 (.72)
Unambiguous feedback	3.94	.64	–.23	–.03	.87	.73–.84 (.80)
Concentration on task	3.66	.69	–.17	.01	.85	.65–.86 (.77)
Sense of control	3.80	.61	–.17	.11	.83	.72–.76 (.75)
Loss self-consciousness	3.36	.85	–.03	–.23	.89	.74–.88 (.83)
Transformation of time	3.49	.79	–.42	.53	.87	.70–.88 (.79)
Autotelic experience	4.20	.61	–.72	.54	.83	.65–.84 (.75)
CFA Model Fit 9 first-order factors: χ^2 = 1380.96, *df* = 558; CFI = .98; NNFI = .98; RMSEA = .05; SRMR = .04						
CFA Model Fit Higher order model: χ^2 = 1603.14, *df* = 585; CFI = .98; NNFI = .97; RMSEA = .05; SRMR = .06						
STATE 36 item						
Challenge skill balance	3.70	.66	–.44	.42	.76	.42–.80 (.68)
Action awareness	3.32	.91	–.22	–.69	.90	.80–.87 (.84)
Clear goals	3.94	.60	–.33	.53	.80	.67–.78 (.72)
Unambiguous feedback	3.85	.63	–.66	1.23	.86	.75–.80 (.78)
Concentration on task	3.69	.81	–.34	–.41	.87	.64–.90 (.79)
Sense of control	3.72	.76	–.52	.14	.88	.73–.86 (.80)
Loss self-consciousness	3.85	.90	–.70	.01	.90	.74–.93 (.84)
Transformation of time	3.50	.83	–.37	.02	.85	.64–.82 (.77)

(continued)

Table 1 ***(continued)***

	Mean/5	*SD*	Skewness	Kurtosis	Reliability	CFA Load Range (mean)
Autotelic experience	4.13	.69	–.81	1.09	.86	.70–.84 (.78)
CFA Model Fit 9 first-order factors: χ^2 = 1332.89, *df* = 558; CFI = .98; NNFI = .97; RMSEA = .05; SRMR = .05						
CFA Model Fit Higher order model: χ^2 = 1717.60, *df* = 585; CFI = .97; NNFI = .96; RMSEA = .06; SRMR = .08						
DISPOSITION 9 item						
Independent short	3.82	.48	–.15	–.01	.77	..30–.69 (.54)
CFA Model Fit: χ^2 = 145.27, *df* = 27; CFI = .95; NNFI = .93; RMSEA = .08; SRMR = .05						
STATE 9 item						
Independent short	3.78	.54	–.50	1.49	.77	.13–.69 (.52)
CFA Model Fit: χ^2 = 462.04, *df* = 27; CFI = .87; NNFI = .83; RMSEA = .14; SRMR = .08						
DISPOSITION 9 item						
Embedded short	3.75	.48	–.15	.37	.77	.25–.73 (.54)
CFA Model Fit: χ^2 = 72.58, *df* = 27; CFI = .98; NNFI = .97; RMSEA = .05; SRMR = .04						
STATE 9 item						
Embedded short	3.73	.51	–.15	.28	.75	.02–.73 (.50)
CFA Model Fit: χ^2 = 183.45, *df* = 27; CFI = .90; NNFI = .87; RMSEA = .11; SRMR = .07						

"Overall, most things I do turn out well "; Cronbach's α = .87; eight items) measures were drawn from the Physical Self Description Questionnaire (PSDQ; Marsh et al., 1994) and were rated on a scale of 1 (*false*) to 6 (*true*). The *perceived competence* measures (e.g., "On a scale of 1–10, how good at your main activity do you consider yourself to be?"; two items) were rated on a scale of 1 (*extremely poor*) to 10 (*extremely good*). The *anxiety* (e.g., "I feel nervous and restless"; Cronbach's α = .83; 20 items) measure was from Spielberger's (1983) Trait Anxiety Scale, rated on a scale of 1 (*almost never*) to 4 (*almost always*). *Positive well-being* (e.g., "I feel great"; Cronbach's α = .85; four items), *psychological distress* (e.g., "I feel awful"; Cronbach's α = .86; four items), and *fatigue* (e.g., "I feel exhausted"; Cronbach's α = .88; four items) were from Subjective Exercise Experience Scale (SEES; McAuley & Courneya, 1994) which uses a list of state-based feelings to assess these constructs, on a rating scale of 1 (*not at all*) to 7 (*very much so*).

Confirmatory Factor Analysis

Confirmatory factor analysis (CFA), performed with LISREL 8.80 (Jöreskog & Sörbom, 2006), was used to test the hypothesized models. Maximum likelihood estimation was employed in these analyses. In evaluating the goodness of fit of alternative models, the comparative fit index (CFI), the nonnormed fit index (NNFI), the root mean square error of approximation (RMSEA), and the standardized root mean residual (SRMR) are emphasized. For the RMSEA, values at or less than .05 and .08 are taken to reflect a close and reasonable fit respectively (see Marsh, Balla, & Hau, 1996). However, where the CFI and NNFI are acceptable and the sample relatively small, the reader is urged to be mindful that larger RMSEA values can over-reject models that are true in the population (Hu & Bentler, 1998). Further, RMSEA is overly sensitive when the number of indicators is small (Kenny & McCoach, 2003), and thus caution is needed when interpreting these values for the short flow scales presented in this study. For the CFI and NNFI, values at or greater than .90 and .95 are typically taken to reflect acceptable and excellent fits to the data respectively (Hu & Bentler, 1998, 1999; McDonald & Marsh, 1990). Hu and Bentler (1998, 1999) recommended the use of SRMR as an absolute fit index, with values close to .08 being regarded as reasonable.

Missing Data

Missing data are a potentially important problem, more so when the amount of missing data exceeds 5% (e.g., Graham & Hoffer, 2000). A growing body of research has emphasized potential problems with traditional pairwise, listwise, and mean substitution approaches to missing data (e.g., Graham & Hoffer, 2000), leading to the implementation of the expectation maximization algorithm, the most widely recommended approach to imputation for missing data, as operationalized using missing value analysis in LISREL. In the current study, less than 5% of the data were missing and so the expectation maximization algorithm was considered an appropriate procedure for all studies.

Multigroup CFA and Tests of Invariance

In assessing the flow scales, it is important to assess for any differences in factor structure and whether a given instrument is equivalent across groups or forms. Testing for factor invariance essentially involves comparing a number of models in which aspects of the factor structure are systematically held invariant across groups/forms and assessing fit indices when elements of these structures are constrained (see Marsh, 1993). Relatively invariant fit indices are indicative of invariant factor structure. The present analyses examined the fit indices for a number of models that held successive elements of the factor structure invariant across samples and forms. We propose two levels at which to assess parameter invariance. The first—and most critical—is invariance in factor loadings, identified as the minimum criterion for establishing invariance (Marsh, 1993). The second level is invariance in correlations, variances, and uniquenesses—with invariance in uniquenesses desirable but not indicative of poor instrumentation in the context of invariance in loadings, correlations, and variances.

Results

Within-Network Validity: Descriptive Statistics and Model Fit

Long Flow (36-Item Scales). In the first set of within-network analyses, the dispositional and state long flow scales (36-item) were examined. Evidence is presented in Table 1 indicating that each of the measures of the nine factors in both scales is reliable and approximately normally distributed. Confirmatory factor analyses of the two instruments show that there was good fit of the hypothesized models to the data for the nine-factor dispositional long flow (χ^2 = 1380.96, *df* =558; CFI = .98; NNFI = .98; SRMR = .04; RMSEA = .05) and the nine-factor state long flow (χ^2 = 1332.89, *df* =558; CFI = .98; NNFI = .97; SRMR = .05; RMSEA = .05). The fit of the higher order model for both dispositional long flow (χ^2 = 1603.14, *df* = 585; CFI = .98; NNFI = .97; SRMR = .06; RMSEA = .05; mean higher order loading = .65) and state long flow (χ^2 = 1717.60, *df* = 585; CFI = .97; NNFI = .96; SRMR = .08; RMSEA = .06; mean higher order loading = .62) was also good and largely mirrored the fit of the nine-factor model.

Short Flow (9-Item Scales). Short flow comprises a subset of nine items from the long (36-item) scale. Two short flow models were examined. The first assessed a set of independent (i.e., stand alone and not part of the 36-item form) nine items. The second assessed the nine items that were part of the long 36-item set (i.e., embedded in the 36-item form). As can be seen in Table 1, the independent dispositional, independent state, embedded dispositional, and embedded state short flow factors are reliable and approximately normally distributed. However, the *dispositional* short flow factors fit the data much better (independent χ^2 = 145.27, *df* = 27; CFI = .95; NNFI = .93; SRMR =.05; RMSEA = .08; embedded χ^2 = 72.58, *df* = 27; CFI = .98; NNFI = .97; SRMR = .04; RMSEA = .05) than the *state* short flow factors (independent χ^2 = 462.04, *df* = 27; CFI = .87; NNFI = .83; SRMR =.08; RMSEA = .13; embedded χ^2 = 183.45, *df* = 27; CFI = .90; NNFI =

.87; SRMR = .07; RMSEA = .11), which did not reach acceptable criterion levels of fit across all indices. This may reflect the fact that more general dispositional ratings of flow do not discriminate so readily between factors and so a short measure pooling factors does not markedly reduce fit. On the other hand, more situation-specific statelike measures may require greater discrimination between factors and a short measure drawing these factors together does reduce relative fit. Indeed, the mean interscale correlation for the 36-item dispositional scale was .43, whereas for the state scale it was .37.

A subset from the total short state sample was examined to further assess possible reasons for the poor fit obtained with the total sample. The largest intact group from this sample was netball/volleyball, and with an *N* of 220, was large enough to run an independent CFA to examine the fit of the model to a situation-specific cohort. The fit of the state short factor for this sample was acceptable (χ^2 = 110.74, *df* = 27; CFI = .93; NNFI = .90; SRMR = .06; RMSEA = .12), other than there being a high RMSEA value, which as we have indicated may be due to sample size and the small number of indicators in the short scale.

Correlations Between Long and Short Scales. Latent correlations between the nine factors comprising the 36-item (long) flow scales and the nine-item (short) flow scales were examined to assess the extent to which the short items captured the essence of their corresponding long factor. Correlations were conducted between matching data, that is, where there was both long and short measures completed by the same respondents. This gave *N*s of 580 for the dispositional scales and 475 for the state scales. There was good fit for the dispositional model (χ^2 = 1660.98, *df* = 801; CFI = .98; NNFI = .98; SRMR = .04; RMSEA = .05) and the state model (χ^2 = 1683.71, *df* = 801; CFI = .98; NNFI = .97; SRMR = .05; RMSEA = .05). Findings revealed that, for both dispositional and state scales, short items correlated at acceptably high levels with their long latent factor counterpart for dispositional (range *r* = .66–.83; mean *r* = .76) and state (range *r* = .65–.82; mean *r* = .73) models. Hence, both dispositional and state short scales provide good representation of their corresponding long versions. In subsequent higher order CFAs, the correlation between the higher order long dispositional factor and the short flow factor was .97 (χ^2 = 4157.35, *df* = 935; CFI = .94; NNFI = .93; SRMR = .08; RMSEA = .07), and the correlation between the higher order long state factor and its short flow factor counterpart was .89 (χ^2 = 4316.60, *df* = 935; CFI = .92; NNFI = .92; SRMR = .10; RMSEA = .08)—also indicating high correspondence between long and short forms.

Between-Network Validity: Relationships With Key Correlates

Between-network correlates for dispositional flow were dispositional intrinsic motivation, perceived competence, sport self-concept, physical self-concept, general self-concept, and anxiety. For the state flow measures, between-network correlates were situational intrinsic motivation, positive well-being, psychological distress, and fatigue. Correlates were chosen from existing valid physical activity measures to match dispositional and state flow measures accordingly.

Long Flow (36-Item Scales)

Table 2 shows that the long flow measures were related to the between-network correlates in hypothesized directions. The long flow scales also provide information about the specific dimensions of flow most related to key correlates. For dispositional flow, key factors were challenge-skill balance, sense of control, clear goals, and autotelic experience. Key correlates were perceived competence, anxiety (negative relationship), and intrinsic motivation. For state flow, autotelic experience was a salient factor, and positive well-being and intrinsic motivation were key correlates.

Short Flow (9-Item Scales)

Findings in Table 2 demonstrate that the dispositional and state flow short scales relate to the between-network correlates in predicted ways, and in a pattern very similar to that of the long flow scales. Interestingly, at an aggregate level the short measures evinced higher between-network correlations than any individual factor in the long 36-item scale—with the dispositional short flow measures yielding the highest between-network correlations. In addition, the independent and embedded short measures reflected quite parallel profiles in that they fit the data in similar ways and related to between-network correlates in similar ways.

Between-Network Validity: Invariance Across Scales

Seven multigroup CFAs were conducted to assess invariance across (a) dispositional and state long 36-item flow, (b) independent dispositional and state short flow, (c) embedded dispositional and state short flow, (d) dispositional independent and embedded short flow, (e) state independent and embedded short flow, (f) dispositional long scale score (aggregating each multi-item factor into a mean score for that factor) and dispositional short flow, and (g) state long scale score (aggregating each multi-item factor into a mean score for that factor) and state short flow.[1] In each of these sets of analyses, five models were assessed. The first allowed all factor loadings, uniquenesses, and correlations/variances to be freely estimated; the second held the factor loadings invariant; the third held factor loadings and uniquenesses invariant; the fourth held the factor loadings and correlations/variances invariant; and the fifth held the factor loadings, the uniquenesses, and the correlations/variances invariant. As indicated earlier, two levels of invariance are of particular interest. The first examines the more critical factor loading invariance. The second examines invariance across other parameters.

Results in Table 3 indicate that when factor loadings were held invariant across models, fit indices did not change markedly. Indeed, the application of recommended criteria for evidence of lack of invariance (i.e., a change of greater than 0.01 in fit indices; see Cheung & Rensvold, 2002) indicated that there was relative invariance across factor loadings. Hence, at the more critical level (i.e., factor loadings), there was support for invariance. Modest invariance also existed at the correlation/variance level. However, for all models except the long 36-item model, there was significant decline in fit at the uniqueness level. Differences in uniquenesses can suggest there may be another method-based (e.g., as a function of item wording or structure) or substantive factor explaining variance in flow for

Table 2 Long (36-Item) and Short (9-Item) Flow Between Network Correlations

	Intrinsic Motivation	Perceived Competence	Sport PSDQ	Physical PSDQ	General PSDQ	Anxiety	Positive Well-Being	Psychological Distress	Fatigue	Mean Absolute *r*
DISPOSITION 36 item										
Challenge skill balance	.48	.69	.53	.42	.42	–.39	—	—	—	.49
Action awareness	.17	.53	.39	.29	.27	–.32	—	—	—	.33
Clear goals	.40	.43	.35	.30	.42	–.38	—	—	—	.38
Unambiguous feedback	.25	.42	.36	.25	.30	–.30	—	—	—	.31
Concentration on task	.48	.34	.35	.29	.27	–.31	—	—	—	.34
Sense of control	.32	.47	.39	.38	.43	–.49	—	—	—	.41
Loss self-consciousness	.18	.09	.09	.33	.28	–.38	—	—	—	.23
Transformation of time	.22	.09	.05	.11	.10	–.03	—	—	—	.15
Autotelic experience	.54	.33	.29	.33	.40	–.37	—	—	—	.38
Mean *r*	.34	.38	.31	.30	.32	–.36	—	—	—	—
CFA Model Fit: $\chi^2 = 9921.32$, $df = 3810$; CFI = .97; NNFI = .97; RMSEA = .05; SRMR = .06										
STATE 36 item										
Challenge skill balance	.24	—	—	—	—	—	.42	–.26	.05	.24
Action awareness	.13	—	—	—	—	—	.17	–.10	.08	.12
Clear goals	.45	—	—	—	—	—	.47	–.30	.01	.31

(continued)

Table 2 *(continued)*

	Intrinsic Motivation	Perceived Competence	Sport PSDQ	Physical PSDQ	General PSDQ	Anxiety	Positive Well-Being	Psychological Distress	Fatigue	Mean Absolute *r*
Unambiguous feedback	.34	—	—	—	—	—	.34	–.23	.01	.23
Concentration on task	.33	—	—	—	—	—	.42	–.26	–.01	.26
Sense of control	.23	—	—	—	—	—	.39	–.33	–.07	.26
Loss self-consciousness	.27	—	—	—	—	—	.38	–.34	–.15	.29
Transformation of time	.34	—	—	—	—	—	.33	–.15	–.08	.23
Autotelic experience	.64	—	—	—	—	—	.82	–.59	–.12	.54
Mean *r*	.33	—	—	—	—	—	.42	–.28	–.03	—
CFA Model Fit: χ^2 = 2500.68, *df* = 1196; CFI = .97; NNFI = .97; RMSEA = .05; SRMR = .05										
DISPOSITION 9 item										
Independent short	.50	.61	.47	.39	.54	–.51	—	—	—	.50
CFA Model Fit: χ^2 = 7282.08, *df* = 1870; CFI = .96; NNFI = .96; RMSEA = .07; SRMR = .06										
STATE 9 item										
Independent short	.41	—	—	—	—	—	.56	–.42	–.07	.37
CFA Model Fit: χ^2 = 1028.95, *df* = 265; CFI = .94; NNFI = .93; RMSEA = .08; SRMR = .08										
DISPOSITION 9 item										
Embedded short	.50	.56	.47	.44	.51	–.51	—	—	—	.50
CFA Model Fit: χ^2 = 5932.69, *df* = 1870; CFI = .97; NNFI = .97; RMSEA = .06; SRMR = .06										
STATE 9 item										
Embedded short	.47	—	—	—	—	—	.64	–.43	–.05	.39
CFA Model Fit: χ^2 = 927.51, *df* = 265; CFI = .95; NNFI = .94; RMSEA = .07; SRMR = .07										

Table 3 Long (36-Item) and Short (9-Item) Flow Invariance Tests

Model—Invariance Across Dispositional and State—36 Item	Chi Square	*df*	CFI	NNFI	RMSEA	SRMR
All parameters are free (no invariance)	2713.85	1116	.98	.98	.05	.05
FIRST-ORDER LOADINGS are invariant	2758.34	1143	.98	.98	.05	.05
FIRST-ORDER LOADINGS and UNIQUENESSES are invariant	3145.97	1179	.97	.97	.05	.06
FIRST-ORDER LOADINGS and CORRELATIONS/VARIANCES are invariant	2990.48	1188	.98	.97	.05	.07
FIRST-ORDER LOADINGS, CORRELATIONS/VARIANCES, and UNIQUENESSES are invariant	3416.50	1224	.97	.97	.06	.09
Model—Invariance Across Independent Dispositional and State—9-Item Short	**Chi Square**	***df***	**CFI**	**NNFI**	**RMSEA**	**SRMR**
All parameters are free (no invariance)	411.81	54	.91	.88	.11	.08
FIRST-ORDER LOADINGS are invariant	428.24	62	.91	.90	.10	.09
FIRST-ORDER LOADINGS and UNIQUENESSES are invariant	655.19	71	.87	.87	.12	.11
FIRST-ORDER LOADINGS and CORRELATIONS/VARIANCES are invariant	434.65	63	.91	.90	.10	.10
FIRST-ORDER LOADINGS, CORRELATIONS/VARIANCES, and UNIQUENESSES are invariant	666.47	72	.87	.87	.12	.14
Model—Invariance Across Embedded (of 36-Item Instrument) Dispositional and State—9-Item Short	**Chi Square**	***df***	**CFI**	**NNFI**	**RMSEA**	**SRMR**
All parameters are free (no invariance)	265.06	54	.95	.93	.08	.07
FIRST-ORDER LOADINGS are invariant	291.08	62	.94	.93	.08	.08
FIRST-ORDER LOADINGS and UNIQUENESSES are invariant	429.65	71	.92	.92	.09	.09
FIRST-ORDER LOADINGS and CORRELATIONS/VARIANCES are invariant	294.26	63	.94	.93	.08	.08

(continued)

Table 3 *(continued)*

Model—Invariance Across Dispositional and State—36 Item	Chi Square	*df*	CFI	NNFI	RMSEA	SRMR
FIRST-ORDER LOADINGS, CORRELATIONS/VARIANCES, and UNIQUENESSES are invariant	436.99	72	.91	.91	.09	.11
Model—Invariance Across Dispositional Independent and Embedded—9-Item Short	**Chi Square**	***df***	**CFI**	**NNFI**	**RMSEA**	**SRMR**
All parameters are free (no invariance)	227.71	54	.96	.95	.07	.05
FIRST-ORDER LOADINGS are invariant	277.07	62	.95	.94	.07	.07
FIRST-ORDER LOADINGS and UNIQUENESSES are invariant	394.07	71	.92	.92	.08	.08
FIRST-ORDER LOADINGS and CORRELATIONS/VARIANCES are invariant	277.43	63	.95	.94	.07	.07
FIRST-ORDER LOADINGS, CORRELATIONS/VARIANCES, and UNIQUENESSES are invariant	394.41	72	.92	.92	.08	.08
Model—Invariance Across State Independent and Embedded—9-Item Short	**Chi Square**	***df***	**CFI**	**NNFI**	**RMSEA**	**SRMR**
All parameters are free (no invariance)	449.15	54	.89	.85	.12	.08
FIRST-ORDER LOADINGS are invariant	553.54	62	.86	.84	.13	.11
FIRST-ORDER LOADINGS and UNIQUENESSES are invariant	607.62	71	.85	.85	.12	.11
FIRST-ORDER LOADINGS and CORRELATIONS/VARIANCES are invariant	553.53	63	.86	.84	.13	.11
FIRST-ORDER LOADINGS, CORRELATIONS/VARIANCES, and UNIQUENESSES are invariant	607.56	72	.85	.85	.12	.11
Model—Invariance Across Scale Score Dispositional Long and 9-Item Dispositional Short	**Chi Square**	***df***	**CFI**	**NNFI**	**RMSEA**	**SRMR**
All parameters are free (no invariance)	346.76	54	.95	.93	.09	.05
FIRST-ORDER LOADINGS are invariant	349.17	62	.95	.94	.08	.05
FIRST-ORDER LOADINGS and UNIQUENESSES are invariant	672.19	71	.89	.89	.11	.09

Table 3 *(continued)*

Model—Invariance Across Dispositional and State—36 Item	**Chi Square**	***df***	**CFI**	**NNFI**	**RMSEA**	**SRMR**
FIRST-ORDER LOADINGS and CORRELATIONS/VARIANCES are invariant	349.64	63	.95	.94	.08	.06
FIRST-ORDER LOADINGS, CORRELATIONS/VARIANCES, and UNIQUENESSES are invariant	675.30	72	.89	.89	.11	.10
Model—Invariance Across Scale Score State Long and 9-Item State Short	**Chi Square**	***df***	**CFI**	**NNFI**	**RMSEA**	**SRMR**
All parameters are free (no invariance)	589.11	54	.88	.84	.14	.08
FIRST-ORDER LOADINGS are invariant	594.55	62	.88	.86	.13	.08
FIRST-ORDER LOADINGS and UNIQUENESSES are invariant	825.23	71	.82	.82	.15	.11
FIRST-ORDER LOADINGS and CORRELATIONS/VARIANCES are invariant	597.83	63	.88	.86	.13	.09
FIRST-ORDER LOADINGS, CORRELATIONS/VARIANCES, and UNIQUENESSES are invariant	832.53	72	.82	.82	.15	.13

one group more than another. However, balancing this against the invariant factor loadings (that we emphasize as the necessary first level of invariance) suggests validity for the fundamental elements of the factor structure across domains. Also to note—and consistent with Table 1 findings—whenever state short measures were included in models, the overall fit declined, particularly when uniquenesses were constrained to be invariant.

Discussion

The present study proposed and assessed an expanded flow measurement framework encompassing the *long* multidimensional aspect of flow and the unidimensional *short* aspect of flow. In general, the flow measures demonstrated acceptable model fit, reliability, and distribution properties (the state short measure received qualified support from a CFA perspective). Correlational analyses between the long and short scales demonstrated that each short item selected to represent its corresponding long latent factor correlated very highly with its long form counterpart, as did the higher order long flow factor with the first-order short factor. From a between-network perspective, the flow measures reflected associations with key correlates in parallel and hypothesized ways, and demonstrated invariance in factor loadings across a diverse sample. We conclude, therefore, that this expanded flow framework and its associated measures are appropriate for continuing empirical research into assessment of flow.

The results of the current study advance earlier flow scale research in several ways. First, the scales received construct validity support in a large and diverse physical activity sample, with sport and exercise activities both well represented. Second, the existing long flow scales and the new short flow scales were examined simultaneously in the one study, allowing for a measure with existing empirical support to be used to benchmark the psychometric properties of a new flow measure. Third, the demonstration of relationships with both previously related and new correlates, as well as the inclusion of invariance testing, substantively extends the between-network validity assessment of the flow scales.

The two measures presented in this study were designed to fulfill different purposes, and thus decisions on which measure to use should be guided by the research purpose and constraints. Drawing on the findings from the current study, we provide some suggested guidelines for uses of the two different types of flow measures.

The Long Flow Scales

The long flow scales provide a rich, multidimensional assessment of the nine dimensions of flow. Where a fine-grained description of flow characteristics according to the theoretical formulation of Csikszentmihalyi (1990) is desired, the long flow scales are the measure of choice. The long scales are also useful for intervention-based studies, where the impact of an intervention across the nine flow components can be assessed. Further, from an applied perspective, the nine-dimensional flow model provides specific and targeted direction for intervention-based research aimed at trying to increase the likelihood of flow occurring. Recent research highlights the fact that the more targeted and differentiated the intervention work, the more effective it is in bringing about important change (e.g., Martin, 2008a).

From a measurement perspective, the long 36-item flow model fit well in both dispositional and statelike contexts. Hence, if researchers are aiming for a rich multidimensional perspective on experience in targeted performance-based activities, the 36-item flow measures perform well. The long flow factors exhibited substantive relationships with key correlates in this study. The autotelic experience dimension exhibited the strongest associations with the set of state correlates, and was also one of the most substantively correlated flow dimensions across the set of dispositional correlates. Sometimes referred to as the *end result* of the other flow factors (Csikszentmihalyi, 1990), autotelic experience captures the essence of being in flow from an experiential perspective. The endorsement of the autotelic experience flow dimension in the current study is consistent with qualitative research (e.g., Jackson, 1996), indicating that this dimension is central to elite athletes' flow experiences.

The CFAs on the long flow scales, in this and previous studies (e.g., Jackson & Eklund, 2002; Jackson & Marsh, 1996), have demonstrated that both the dispositional and state versions are valid measures. Hence, choice of instrument devolves to applied considerations such as that of the type of context in question. The context-general dispositional version and the context-specific state version provide psychometrically sound tools for assessing the nine-dimensional characterization of flow. In addition, the fit of the higher order model for long flow was shown in this study to be acceptable, and there are occasions when a higher order measure is useful in research settings. For example, in predictive models and/or when flow is not a central focus, a higher order measure of the flow construct may be deemed more functional than the multidimensional first-order measure—the fit indices indicated that use of a higher order structure in such situations would not unacceptably diminish the overall fit of any given model including a higher order flow factor.

The Short Flow Scales

The short flow scales provide a brief assessment of the nine dimensional conceptualization of flow. It is a practical tool for situations where constraints prevent the use of the more complete long measure—and where an aggregate flow score that encompasses the nine dimensions is of interest. Thus, the short scales may be of interest in research such as multitrait multimethod studies where a large number of constructs are being simultaneously assessed, as well as in self–other agreement studies. The short flow scale items capture an aggregate profile of flow characteristics in a simple, generic format, making them relevant to a range of settings.

From a measurement perspective, dispositional short flow fit well, and much better than state flow. Thus, from a psychometric viewpoint, the dispositional short measure is a suitable tool for providing a brief assessment of flow from the nine-dimensional conceptualization. Further research is needed to assess the state short measure. It may be that a more generalized rating of flow contained within the dispositional short measure does not discriminate as readily between factors and thus pooling factors into the short measure does not reduce fit. However, the more situation-specific state short measure likely requires more discrimination between factors and so when they are drawn together into a short measure, fit is reduced. The relatively lower interscale correlation for the state short measure compared with the dispositional short measure in this study supports this idea. Further, when the context is held constant, the state short scale fit improves, as we demonstrated with the fit of this model in a specific sport context sample. The substantial associations with a range of key correlates in this study in parallel and hypothesized ways further supports the convergent and divergent validity of the short scales.

Limitations and Future Directions

As argued in previous research (e.g., Csikszentmihalyi, 1992; Jackson & Eklund, 2002; Jackson & Marsh, 1996), all empirical assessments of flow are only partial reflections of the experience. As is the case with all experiential phenomena, flow cannot *easily* be quantified by psychometrics or fully illuminated through investigative interviewing. No single measurement approach will provide trouble-free assessments of the flow experience. The approaches presented in this research provide tools to tap into aspects of the flow experience. Together, the two scales presented herein allow for measurement of dimensions of the flow experience, and for examination of differences in the occurrence of flow across individuals or settings. A multimethod approach to researching flow will result in improved understanding of flow, and how to achieve it.

The data presented in this study are subject to the limitations of self-report. Although a logical and defensible methodology given the substantive focus, it is important to also conduct research examining flow using data derived from additional sources. Subjective sources could include coaches, teachers, parents, and peers. "Objective" sources could include achievement and outcome data relevant to the performance domain. Further limitations of self-report instruments of the type assessed in the current study include those emanating from retrospective recall, and from group average representations of experience. It is acknowledged that the information provided by the scales presents a partial picture of what the experience of being in flow is like.

The short scale is a relatively new instrument, although initial research has shown it to be a promising tool in situations where flow is one of several constructs being assessed—one of the purposes for which the short versions were developed (Martin et al., 2006; Martin, in press, 2008b). Research is needed to further examine validity and reliability of the short forms. The state short scale, in particular, requires additional research to examine its psychometric properties, and we recommend that this be done across a diversity of domains. Although some CFA fit indices for the combined sample did not reach criterion levels, other psychometric data presented in this study—fit indices for a context-specific sample, support for hypothesized associations with key correlates, correlations between short and long forms, and internal consistency estimates—indicate that the short state scale has empirical potential. Because of the strong conceptual model from which both the long and the short flow scales were developed, we believe it is important to continue to assess the empirical properties of the nine-item short forms, and to do so across domains, before drawing any firm conclusions regarding the utility of these measurement instruments.

Historically, the flow scales have primarily been used in sport and other physical activity settings. However, research has demonstrated the utility of the long flow scales in other settings, including theater (Martin & Cutler, 2002), Web-based instructional activity (Chan & Repman, 1999), yoga (Penman, Cohen, Stevens, & Jackson, 2008), music (Jackson & Eklund, 2004b), and other creative or performing arts (Jackson & Eklund, 2004b). Further, the dispositional short flow scale has been assessed in physical activity (Martin et al., 2006), and more recently, comprehensively in work, sport, and music (Martin & Jackson, 2008; Martin, 2008b, in press). The long and, in particular, the short scales are adaptable to a variety of settings. Future research could use the scales to examine commonalities and differences in self-reported flow experience across different contexts, as well as across cultures.

The long scales provide a tool for examining the relative incidence and influence of the nine flow dimensions on the experience of flow across individuals and contexts. They also are useful for intervention research, where flow may be a predictor, mediator, or outcome variable, and for continuing to examine relationships between flow and other constructs. In addition, longitudinal and intervention research would be useful in examining the extent to which flow can change and, if so, how to most effectively bring about such change. Use of the dispositional version of the long scale in different contexts may help with the identified need of increasing understanding of the autotelic personality and what qualities are associated with people who are skilled at experiencing flow in their lives (Nakamura & Csikszentmihalyi, 2002).

Both flow measures also provide opportunities for exploring the extent to which group setting such as teams, classrooms, and workplaces are operating to facilitate optimal experience. Using multilevel modeling (e.g., Goldstein, 2003), it is possible to assess the relative influence of individual- and group-level factors. Future research could thus explore the influence of group-level flow climates relative to individual-level variation in flow. Another technique that would be a useful tool for future research would be multitrait multimethod analyses (Campbell & Fiske, 1959), where scales representing a diverse range of constructs could be compared with flow responses obtained both by participants and by secondary sources.

Conclusion

The present analyses of the long flow scales have provided further support indicating that both the dispositional and state versions are valid and reliable measures. Analyses of the new short flow scale have shown the dispositional version to be a valid and reliable measure. Further research is needed to examine the properties of the state version of the short scale, with the current study providing limited validity support from a confirmatory factor analytic perspective. Psychometric support for the state short scale was, however, observed in internal consistency and external validity in terms of associations with key psychological constructs. Future research should examine the short flow scales across contexts and levels of participant experience to provide further assessment of its psychometric properties. Analyses also confirmed between-network validity via relative invariance across different scales and models, as well as via correlations with cognate constructs. Together, the scales assessed in this study provide useful tools for diverse research goals. They are also relevant to practitioners interested in enhancing outcomes that are related to being in flow.

Note

1. For completeness, invariance tests on all four flow forms were also conducted across gender (male/female) and competitive status (competitive/noncompetitive). Invariance tests were not conducted on type of physical/sporting activity as group cell sizes were too small for multi-group CFAs—this is an area for future research. Being noncentral subsidiary analyses, only the more widely used fit indices (CFI, RMSEA–CFI, RMSEA; see McDonald & Marsh, 1990) and only the least and most restrictive solutions are presented here. As with central analyses, a change of greater than .01 is evidence of a lack of invariance (Cheung & Rensvold, 2002). Results indicated broad invariance for gender: standard dispositional flow (all parameters free: CFI = .98, RMSEA = .05 vs all parameters invariant: CFI = .97, RMSEA = .06), composite dispositional flow (all parameters free: CFI = .95, RMSEA = .08 vs all parameters invariant: CFI = .95, RMSEA = .08), standard state flow (all parameters free: CFI = .96, RMSEA = .06 vs all parameters invariant: CFI = .96, RMSEA = .07), composite state flow (all parameters free: CFI = .89, RMSEA = .13 vs all parameters invariant: CFI = .88, RMSEA = .12—but note the lower fit more generally for composite state flow). On all but one flow form, results also indicated broad invariance for competitive status: standard dispositional flow (all parameters free: CFI = .98, RMSEA = .05 vs all parameters invariant: CFI = .97, RMSEA = .06), composite dispositional flow (all parameters free: CFI = .94, RMSEA = .09 vs all parameters invariant: CFI = .88, RMSEA = .12—with lack of invariance here due to uniquenesses), standard state flow (all parameters free: CFI = .96, RMSEA = .06 vs all parameters invariant: CFI = .96, RMSEA = .07), composite state flow (all parameters free: CFI = .90, RMSEA = .12 vs all parameters invariant: CFI = .88, RMSEA = .12—again, however, note the lower fit more generally for composite state flow).

Acknowledgments

This article was in part prepared while the first author was a recipient of a UQ Re-Entry Fellowship for women, and while the second author was Visiting Research Fellow at Oxford University. Data reported here for the DFS-2 and FSS-2 were also used for cross-cultural comparison purposes in Kawabata (2008), and a portion of the DFS-2 data were also used for criterion-validity purposes in Mallett, Kawabata, Newcombe, Otero-Forero, & Jackson (2007). The authors would like to acknowledge Kawabata, Mallett, and colleagues for their contribution to the data reported herein.

References

Amiot, C.E., Gaudreau, P., & Blanchard, C. (2004). Self-determination, coping, and goal attainment in sport. *Journal of Sport & Exercise Psychology, 26,* 396–411.

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

Chan, T.S., & Repman, J. (1999). Flow in web based instructional activity: An exploratory research project. *International Journal of Educational Telecommunications, 5,* 225–237.

Cheung, G.W., & Rensvold, R.B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9,* 233–255.

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety*. San Francisco: Jossey-Bass.
Csikszentmihalyi, M. (1988). Introduction to a theoretical model of optimal experience. In Csikszentmihalyi, M., & Csikszentmihalyi, I. (Eds.), (1988). *Optimal experience: Psychological studies of flow in consciousness* (pp. 3–14). Cambridge, MA: Cambridge University Press.
Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper & Row.
Csikszentmihalyi, M. (1992). A response to the Kimiecik & Stein and Jackson papers. *Journal of Applied Sport Psychology, 4,* 181–183.
Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the Experience Sampling Method. *The Journal of Nervous and Mental Disease, 175,* 526–536.
Deci, E.L., & Ryan, R.M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.
Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London: Hodder Arnold.
Graham, J.W., & Hoffer, S.M. (2000). Multiple imputation in multivariate research. In T.D. Little, K.U. Schnable, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 201–218). NJ: Erlbaum.
Guay, F., Vallerand, R.J., & Blanchard, C. (2000). On the assessment of situational intrinsic and extrinsic motivation: the Situational Motivation Scale. *Motivation and Emotion, 24,* 175–213.
Hu, L., & Bentler, P.M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3,* 424–453.
Hu, L., & Bentler, P.M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.
Jackson, S.A. (1992). Athletes in flow: A qualitative investigation of flow states in elite figure skaters. *Journal of Applied Sport Psychology, 4,* 161–180.
Jackson, S.A. (1996). Toward a conceptual understanding of the flow experience in elite athletes. *Research Quarterly for Exercise and Sport, 67,* 76–90.
Jackson, S.A., & Csikszentmihalyi, M. (1999). *Flow in sports: The keys to optimal experiences and performances*. Champaign, IL: Human Kinetics.
Jackson, S.A., & Eklund, R.C. (2002). Assessing flow in physical activity: The Flow State Scale-2 (FSS-2) and Dispositional Flow Scale-2 (DFS-2). *Journal of Sport & Exercise Psychology, 24,* 133–150.
Jackson, S.A., & Eklund, R.C. (2004a). *The flow scales manual*. Morgantown, WV: Fitness Information Technology.
Jackson, S.A., & Eklund, R.C. (2004b). Relationships between quality of experience and participation in diverse performance settings. *Australian Journal of Psychology, 56*(Supplement), 193.
Jackson, S.A., Kimiecik, J., Ford, S., & Marsh, H.W. (1998). Psychological corrrelates of flow in sport. *Journal of Sport & Exercise Psychology, 20,* 358–378.
Jackson, S.A., & Marsh, H.W. (1996). Development and validation of a scale to measure optimal experience: The flow state scale. *Journal of Sport & Exercise Psychology, 18,* 17–35.
Jackson, S.A., Thomas, P.R., Marsh, H.W., & Smethurst, C.J. (2001). Relationships between flow, self-concept, psychological skills, and performance. *Journal of Applied Sport Psychology, 13,* 129–153.
Jöreskog, K.G., & Sörbom, D. (2006). *LISREL 8.80*. Chicago: Scientific Software International.
Kawabata, M. (2008). *Optimal experience in physical activity: Examining the multidimensionality of flow across cultures*. Unpublished Doctoral Thesis, University of Queensland, Brisbane, Australia.

Kenny, D.A., & McCoach, D.B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling, 10,* 333–351.

Mallett, C., Kawabata, M., Newcombe, P., Otero-Forero, A., & Jackson, S.A. (2007). Sport Motivation Scale-6 (SMS-6): A revised six factor sport motivation scale. *Psychology of Sport and Exercise, 8,* 600–614.

Marsh, H.W. (1993). The multidimensional structure of academic self-concept: Invariance over gender and age. *American Educational Research Journal, 30,* 841–860.

Marsh, H.W. (1997). The measurement of physical self-concept: A construct validation approach. In K. Fox (Ed.), *The physical self-concept: From motivation to well-being* (pp. 27–58). Champaign, IL: Human Kinetics.

Marsh, H.W., Balla, J.R., & Hau, K.T. (1996). An evaluation of incremental fit indices: A clarification of mathematical and empirical processes. In G.A. Marcoulides & R.E. Schumacker (Eds.), *Advanced structural equation modeling techniques* (pp. 315–353). Hillsdale, NJ: Erlbaum.

Marsh, H.W., Ellis, L.A., Parada, R.H., Richards, G., & Heubeck, B.G. (2005). A short version of the Self Description Questionnaire II: Operationalizing criteria for short-form evaluation with new applications of confirmatory factor analyses. *Psychological Assessment, 17,* 81–102.

Marsh, H.W., & Jackson, S.A. (1999). Flow experiences in sport: Construct validation of multidimensional hierarchical state and trait responses. *Structural Equation Modeling, 6,* 343–371.

Marsh, H.W., Martin, A.J., & Hau, K.T. (2006). A multiple method perspective on self-concept research in educational psychology: A construct validity approach. In M. Eid & E. Diener (Eds). *Handbook of multimethod measurement in psychology*. Washington, DC: American Psychological Association Press.

Marsh, H.W., Richards, G.E., Johnson, S., Roche, L., & Tremayne, P. (1994). Physical Self Description Questionnaire: Psychometric properties and a multitrait-multimethod analysis of relations to existing instruments. *Journal of Sport & Exercise Psychology, 16,* 270–305.

Martin, A.J. (in press).How domain specific are motivation and engagement across school, sport, and music? A substantive-methodological synergy assessing young sportspeople and musicians. *Contemporary Educational Psychology*.

Martin, A.J. (2008a). Enhancing student motivation and engagement: The effects of a multidimensional intervention. *Contemporary Educational Psychology, 33,* 239–269.

Martin, A.J. (2008b). Motivation and engagement in music and sport: Testing a multidimensional framework in diverse performance settings. *Journal of Personality*.

Martin, J.J., & Cutler, K. (2002). An exploratory study of flow and motivation in theater actors. *Journal of Applied Sport Psychology, 14,* 344–352.

Martin, A.J., & Jackson, S.A. (in press). Brief approaches to assessing task absorption and enhanced subjective experience: Examining 'short' and 'core' flow in diverse performance domains. *Motivation and Emotion.*

Martin, A.J., Tipler, D.V., Marsh, H.W., Richards, G.E., & Williams, M.R. (2006). Assessing multidimensional physical activity motivation: A construct validity study of high-school students. *Journal of Sport & Exercise Psychology, 28,* 171–192.

McAuley, E., & Courneya, K.S. (1994). The subjective exercise experience scale: Development and preliminary validation. *Journal of Sport & Exercise Psychology, 16,* 163–177.

McDonald, R.P., & Marsh, H.W. (1990). Choosing a multivariate model: Noncentrality and goodness-of-fit. *Psychological Bulletin, 107,* 247–255.

Nakamura, J., & Csikszentmihalyi, M. (2002). The concept of flow. In C.R. Snyder & S.J. Lopez (Eds.), *Handbook of positive psychology* (pp. 89–105). Oxford: Oxford University Press.

Pelletier, L.G., Fortier, M.S., Vallerand, R.J., Tuson, K.M., Brière, N.M., & Blais, M.R. (1995). Toward a new measure of intrinsic motivation, extrinsic motivation, and amotivation in sports: The Sport Motivation Scale (SMS). *Journal of Sport & Exercise Psychology, 17,* 35–53.

Pelletier, L.G., Fortier, M.S., Vallerand, R.J., & Briere, N.M. (2001). Associations among perceived autonomy support, forms of self-regulation, and persistence: A prospective study. *Motivation and Emotion, 25,* 279–306.

Penman, S., Cohen, M., Stevens, P., & Jackson, S. (2008). *Yoga in Australia: Results of a National Survey*. Retrieved April 11, 2008 from www.YogainAustralia.com

Seligman, M.E.P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *The American Psychologist, 55,* 5–14.

Spielberger, C.D. (1983). *Manual for the State-Trait Anxiety Inventory (Form Y)*. Palo Alto, CA: Consulting Psychologists Press.

Williams, L.J., Ford, L.R., & Nguyen, N. (2002). Basic and advanced measurement models for confirmatory factor analysis. In S.G. Rogelberg (Ed.), *Handbook of research methods in industrial and organizational psychology* (pp. 366–389). Oxford, UK: Blackwell.

Manuscript submitted: December 15, 2007
Revision accepted: May 21, 2008