JOURNAL OF SPORT & EXERCISE PSYCHOLOGY, 2002, **24**, 133-150

Assessing Flow in Physical Activity: The Flow State Scale–2 and Dispositional Flow Scale–2

Susan A. Jackson
Queensland University of Technology

Robert C. Eklund
The University of Western Australia

The Flow State Scale–2 (FSS–2) and Dispositional Flow Scale–2 (DFS–2) are presented as two self-report instruments designed to assess flow experiences in physical activity. Item modifications were made to the original versions of these scales in order to improve the measurement of some of the flow dimensions. Confirmatory factor analyses of an item identification and a cross-validation sample demonstrated a good fit of the new scales. There was support for both a 9-first-order factor model and a higher order model with a global flow factor. The item identification sample yielded mean item loadings on the first-order factor of .78 for the FSS–2 and .77 for the DFS–2. Reliability estimates ranged from .80 to .90 for the FSS–2, and .81 to .90 for the DFS–2. In the cross-validation sample, mean item loadings on the first-order factor were .80 for the FSS–2, and .73 for the DFS–2. Reliability estimates ranged between .80 to .92 for the FSS–2 and .78 to .86 for the DFS–2. The scales are presented as ways of assessing flow experienced within a particular event (FSS–2) or the frequency of flow experiences in chosen physical activity in general (DFS–2).

Key Words: flow research, scale development, construct validity

Interest in the study of flow experiences in physical activity has grown as sport and exercise psychology has come to recognize the importance of the positive side of activity experiences. Flow, as an optimal psychological state, represents those moments when everything comes together for the performer; it is often associated with high levels of performance and a very positive experience. Although flow has been a construct of interest to sport psychology researchers and practitioners for some time, the pursuit of research in this area has been greeted with some hesitation. Some likely reasons include the difficulty in measuring an experiential state as well as uncertainty about how to define, in empirical terms, the flow construct.

S.A. Jackson was with Queensland Univ. of Technology, Australia, at the time of this study; address correspondence to <susan.jackson@bigpond.com> R.C. Eklund is with the Dept. of Human Movement and Exercise Science, Univ. of Western Australia, Perth, Australia.

Research by Jackson (1992, 1995, 1996; Jackson, Kimiecik, Ford, & Marsh, 1998; Jackson & Marsh, 1996; Jackson, Thomas, Marsh, & Smethurst, 2001) has focused on understanding and examining the flow state in physical activity. Beginning with qualitative approaches (Jackson, 1992, 1995, 1996), Jackson explored elite performers' perceptions of flow and how they attained this state during their performances. In an effort to continue examining flow in physical activity, and specifically to understand the relationship of flow to other psychological factors, Jackson (Jackson & Marsh, 1996; Jackson et al., 1998) developed self-report instruments to assess flow experiences. The Flow State Scale (FSS; Jackson & Marsh, 1996) and Dispositional[1] Flow Scale (Jackson et al., 1998) were designed to assess, respectively, flow experiences within a particular event and the dispositional tendency to experience flow in physical activity. These scales were theoretically grounded in Csikszentmihalyi's (1990) nine-dimensional concept of flow. Items were developed from analysis of definitions of the nine flow dimensions, qualitative reports by athletes (Jackson, 1996), and other measures of flow or related experiences (Begly, 1979; Csikszentmihalyi & Csikszentmihalyi, 1988; Privette, 1984; Privette & Bundrick, 1991).

The nine flow dimensions posited by Csikszentmihalyi (1990) are challenge–skill balance, action–awareness merging, clear goals, unambiguous feedback, concentration on task, sense of control, loss of self-consciousness, time transformation, and autotelic experience. Together they represent the optimal psychological state of flow; singly they signify conceptual elements of this state.

Although defining flow as a multidimensional construct comprising the nine dimensions described above, Csikszentmihalyi (e.g., Cskiszentmihalyi & Csikszentmihalyi, 1988) has primarily relied on the challenge–skill balance dimension to measure flow. When challenges and skills are both at personally high levels, flow is predicted to occur. Other experiential dimensions such as anxiety, apathy, and boredom are also predicted via the challenge–skill ratio. This approach of measuring challenges and skills forms the core of the experience sampling method (ESM; Csikszentmihalyi & Larson, 1987), whereby multiple assessments of experience are recorded. While the ESM has proven to be a popular method of assessing flow in daily life, it has yet to be embraced in physical activity settings, where researchers are often interested in discrete assessments and where disrupting performance during the activity is another obstacle to using the ESM approach.

In developing the FSS and DFS for the physical activity setting, it was considered important to have self-report measures that could be used without disrupting performance. However, a more important goal was to develop multidimensional measures that would reflect all nine flow dimensions rather than focusing primarily on the challenge-skill balance ratio. Because the flow construct is multidimensional, measures of flow that tap into all of the flow dimensions should provide a more complete assessment than unidimensional measures.

Taking a multidimensional approach to the measurement of flow, Jackson and Marsh (1996) assessed the utility of a 36-item self-report instrument containing 4 items for each flow dimension. Confirmatory factor analyses demonstrated a satisfactory fit of both a 9-first-order factor model and one higher order model

[1]As noted in Jackson et al. (2001), the Trait Flow Scale was renamed to Dispositional Flow Scale to more accurately represent what it purports to measure.

with a global flow factor. Parameter estimates provided good support for the 9-factor structure with freely estimated factor correlations. The factor loadings ranged from .56 to .88, with a median loading of .74. Correlations between the nine factors showed that although the relationships were all positive, the size of the correlations, varying from .18 to .72 (median r = .50), supported the separation into nine flow factors.

The higher order model presented by Jackson and Marsh (1996) was supportive of a global flow factor, but caution was recommended in using a single factor or score from the FSS to represent flow. Although each of the nine factors loaded on the higher order factor, there was considerable variability in the size of these loadings, ranging from a low of .39 for time transformation to a high of .91 for sense of control. Further, the residual variance estimates for each first-order factor ranged from .24 to .88, indicating that between 24 to 88% of the variance in the first-order factors was unexplained by the higher order factor.

The FSS was developed to assess events recently experienced, thus the instructions to respondents are worded accordingly. Initial FSS data, however, were based on the retrospective recall of physical activity participants reporting on a past flow experience that stood out for them (Jackson & Marsh, 1996). Although useful in generating responses that could reasonably be assumed to emanate from genuine flow experiences, this retrospective approach was a design limitation in that the responses could have been influenced by the passing of time. Subsequent research by Jackson and colleagues (e.g., Jackson et al., 1998; 2001) has captured postevent flow responses with the FSS and provided additional information as to the validity and reliability of the scale.

The DFS was developed subsequent to the FSS, using instructions that focus on the frequency of the experience of flow in order to assess individual differences in the propensity to experience flow. This DFS variation of the FSS was developed because Csikszentmihalyi and others (see Csikszentmihalyi & Csikszentmihalyi, 1988) have proposed that there are individual differences in the ability to experience flow. Certain types of people, Csikszentmihalyi has suggested, may be better psychologically equipped to experience flow, regardless of the situation. The term *autotelic personality* applies to this propensity to experience flow. Just what makes up the autotelic personality is somewhat of a mystery, although several important factors have been identified such as desire for challenge (Logan, 1988) and superior concentration skills (Hamilton, as cited in Csikszentmihalyi & Csikszentmihalyi, 1988). In exploring the constituents of the autotelic personality, Jackson and colleagues (1998; 2001) have identified psychological factors expected to relate to dispositional flow.

In any event, Marsh (1997, 1998) has characterized the establishing of construct validity for an instrument as a multistep process that begins with analysis of factor structure or dimensionality (within-network approach) and moves on to analysis of patterns of relationship between the construct and other constructs (between-network approach). In addition to reporting the results of within-network analyses of FSS and the DFS measurement (Jackson & Marsh, 1996; Marsh & Jackson, 1999), Jackson et al. (1998; 2001) also reported between-network construct validity results. Specifically, theoretically expected patterns of relationship between flow and the psychological constructs logically related to flow were observed between flow and perceived ability, anxiety, and intrinsic motivation (Jackson et al., 1998). Relationships between flow and dimensions of athletic self-concept, as well as with athletes' use of psychological skills, have also been reported (Jackson et al., 2001). In both studies, dispositional flow demonstrated stronger relationships with the various psychological constructs than did the state flow measures; this was an expected finding, given that all of the non-flow constructs were also assessed at a dispositional level.

The relationships between FSS flow ratings and performance correlates—perceived skill, perceived success, subjective performance ratings, overall finishing position—have also been examined. Specifically, flow state dimensions have been reported as being positively correlated with measures of perceived skill and perceived success (Jackson et al., 1998), subjective performance ratings, and overall finishing position (Jackson et al., 2001). Interestingly, performance in both studies was more strongly related to FSS measures than to DFS measures. Together these studies indicate that the FSS and DFS exhibit construct validity and utility for measuring both flow state experiences and the dispositional tendency to experience this state.

Marsh and Jackson (1999) reported a series of sophisticated confirmatory factor analyses to individually and simultaneously evaluate the measurement of FSS and DFS. Overall, good support was presented for the construct validity of the state and dispositional measures. Item loadings on first-order factors ranged from .43 to .89 for the FSS (mean = .78) and from .29 to .86 for the DFS (mean = .74). As in all investigations employing the flow scales (e.g., Jackson & Marsh, 1996; Jackson et al., 2001), internal consistency estimates for both scales were reasonable, varying from .72 to .91 (mean α = .85) for the FSS and from .70 to .88 (mean α = .82) for the DFS.

Not surprisingly, Marsh and Jackson (1999) found that models involving exclusively first-order factors fit marginally better than higher-order factor models. Higher-order factor loadings ranged from .00 to .88 for the FSS (mean = .55) and from .04 to .89 for the DFS (mean = .62). Only one factor had a loading of less than .40 on the higher order factor. The troublesome factor was time transformation; it exhibited essentially no relationship with the global factor in either DFS or FSS measurement in this sample.

The only other published confirmatory factor analytic study of the flow scales that we are aware of, drawing conclusions on the FSS inconsistent with those reported by Jackson and colleagues, was conducted by Vlachopoulos, Karageorghis, and Terry (2000). Using the responses of aerobic dance participants, they concluded that the 9-factor model and the hierarchical model did not show an adequate fit to the data. The fit values obtained by Vlachopoulos et al. were slightly lower than those previously obtained by Jackson and Marsh (1996) and Marsh and Jackson (1999), and may be open to some interpretational debate given the number of items in the flow scales. For example, Marsh (2000) has indicated that it may be difficult for scales with numerous factors and items to reach Hu and Bentler's (1999) revised standards. Nonetheless, the Vlachopoulos et al. findings did highlight areas of weakness consistent with those highlighted by Jackson and colleagues (Jackson & Marsh, 1996; Jackson et al., 2001; Marsh & Jackson, 1999). These relate to the relatively weaker associations between the dimensions, i.e., loss of self-consciousness and time transformation, with the remaining flow dimensions and the global flow factor in the hierarchical factor analytic model.

The analyses of data collected with the original flow scales indicate that, while they perform reasonably well on the whole, there are areas that could be improved. Parameter estimates (Jackson & Marsh, 1996; Marsh & Jackson, 1999; Vlachopoulos et al., 2000) indicate that a small number of items warrant some additional conceptual and empirical consideration. For the most part, first-order flow factors appear to load well on the global flow factor. Nonetheless, the dimensions of loss of self-consciousness and time transformation have not consistently shown satisfactory loadings on the higher-order flow factor.

Appropriate evaluation of the measurement qualities of the flow scales requires consideration of both conceptual and statistical issues. As part of the conceptual evaluation, we obtained feedback on items in the original scale from the developer of the flow model, Csikszentmihalyi (1975, 1990). Several issues were identified as meriting consideration. For example, while none of the sense-of-control items demonstrated statistical weakness, M. Csikszentmihalyi (personal communication, April 21, 1997) noted there was too much emphasis on "total control." He suggested that the measurement domain should focus more on having a *sense* of control (or lack of worry about control) over what one was doing. We took this conceptual feedback into consideration when designing potential replacement items for the scales.

In terms of statistical weakness, the poorest performing item across several analyses has been the second one from the loss-of-self-consciousness subscale. The loadings of this item have ranged from .29 on the DFS and .43 on the FSS (Marsh & Jackson, 1999) to .56 on the FSS (Jackson & Marsh, 1996). This item (i.e., DFS version: "I am not worried about my performance during the event") focuses on a lack of concern for self-presentation or evaluation by others. While this item was designed to tap into a lack of concern or worry for oneself during performance, it appears to be somewhat ambiguous and open to interpretation in different ways.

The first unambiguous feedback item (FSS version: "It was really clear to me that I was doing well") has also shown less than optimal performance, at least in the state version of the inventory. Specifically, Marsh and Jackson (1999) reported that the LISREL modification indexes suggested that the item was substantially related to the state challenge–skill factor. The corresponding dispositional version of this item did not exhibit similar problems. The slight differences in wording between the measurement to tap into the experiential tendency instead of the experience of the state may have resulted in less ambiguity for respondents, and hence produced the different statistical outcomes.

One troubling issue, both conceptually and empirically, is that the time-transformation dimension has consistently loaded weakly on the global flow factor in state measurement of the flow experience (Jackson & Marsh, 1996; Marsh & Jackson, 1999; Vlachopoulos et al., 2000). M. Csikszentmihalyi (personal communication, April 21, 1997) suggested that the wording of some time-transformation items may have been part of the problem. This dimension of flow relates to a relatively unusual perception of the passage of time while in a flow state in that time both seems to slow down and speed up. Csikszentmihalyi suggested that the more common occurrence in flow is a perceptual shortening rather than lengthening of time. Since two of the original items focused on time lengthening while none tapped into time shortening, the original measurement of this dimension may have been inadequate, particularly in assessing the state experience. Additional items were developed as possible alternatives to the original time-transformation items in order to assess whether changing the focus of the scale would improve its performance.

The loss-of-self-consciousness scale has also consistently demonstrated relatively low loadings on the higher order factor (Jackson & Marsh, 1996; Marsh & Jackson, 1999; Vlachopoulos et al., 2000). The reasons for this may include the previously discussed poor fitting item in this subscale, an inadequacy on the part of the remaining items in tapping into the dimension, and/or the lesser relevance of this dimension of flow to athletes and physical activity participants. To address the possibility of the item set not tapping into the dimension well, new items were developed and assessed in the current project.

In summary, the purpose of the present study was to assess the relative usefulness, from a psychometric perspective, of a small pool of new items for the flow scales. Conceptual and statistical issues were considered in developing possible replacements for original items from the FSS and DFS. First, the conceptual adequacy of FSS and DFS measurement was evaluated against extant literature describing the flow experience, and issues were discussed with M. Csikszentmihalyi (personal communication, April 21, 1997). Second, empirical issues on the psychometric performance of the scale, its subscales, and items were examined to ensure that statistical weaknesses in items or higher-order factor loadings were considered along with conceptual issues. Conceptually coherent item replacements were developed and considered. This report discusses confirmatory factor analyses conducted to evaluate and select among these items for new versions of the FSS and DFS.

STUDY 1

Method

Participants

A total of 597 respondents contributed data to the Study 1 analyses, including 391 who provided FSS data and 386 who contributed DFS data (180 provided both FSS and DFS data). These participants ranged in age from 17 to 72 years (M = 26.3, SD = 10.0, ~1% nonresponse). The gender breakdown was 49% males and 51% females (~1% nonresponse).

To be eligible for the study, respondents had to take part in physical activity on a regular basis at least twice a week. Participants engaged in a wide variety of physical activities. There were 33 activity types involving more than a single respondent (activities with only one respondent were collated into an "other" category). These activities ranged from highly competitive sports such as rugby to exercise activities such as weight training. The 10 most common activities reported were touch football (n = 145), triathlon (n = 105), running (n = 65), duathlon (n = 56), surf boat rowing (n = 45), track & field (n = 41), swimming (n = 27), rugby (n = 25), soccer (n = 24), and volleyball (n = 23). A wide range in participation levels was reported, including (but not limited to) international (10%), national (15%), state (24%), and club or school (26%) involvement.

Instruments

Dispositional Measures. The first questionnaire included background demographic information and a version of the Dispositional Flow Scale. This version included the original 36-item DFS along with 13 additional items devised as potential replacements to address the earlier identified conceptual or statistical concerns. Examples of new items included, "I am not concerned with how others may be evaluating me" (loss of self-consciousness) and "I have a sense of control over what I am doing" (sense of control). Respondents were instructed to think about how often they experienced the flow characteristic identified in each item and to estimate that frequency on a 5-point Likert scale ranging from 1 = "never" to 5 = "always".

State Measures. The second questionnaire contained questions about a just-completed activity along with a version of the Flow State Scale. The activity related questions were designed to ground the respondent in the experience of the event to which he or she was responding, with questions about the date, time of activity, time of questionnaire completion, and type of activity, as well as an open-ended description of his or her experience in the event. In addition, there were questions about how similar or different this event was from other times when the respondent took part in the activity, and he or she was asked to rate the challenges, skills, and his or her level of performance in the activity.

The version of the FSS employed in Study 1 was composed of 49 items that included the original 36-item FSS and 13 state measurement versions of the potential replacement items added to the DFS. Examples of new items included, "It was really clear to me how my performance was going" (unambiguous feedback), and "It felt like time went by quickly" (time transformation). Respondents were asked to indicate the extent of their agreement with the items as characterizing their experience in the just-completed event. Responses were on a 5-point Likert scale ranging from 1 = "strongly disagree to 5 = "strongly agree."

Procedures

Participants were recruited from university classes (Human Movement, Psychology), sports teams, and at events such as triathlons. Standard introduction and information instructions were given to all participants, and informed-consent procedures were adhered to. The dispositional measures were completed at a time separate from participants' participation in physical activity. A trained research assistant provided general instructions for completing the questionnaire. Specific instructions were included at the top of the questionnaire.

The revised FSS was given to respondents to complete after participating in their main activity. Questionnaires were either distributed at the conclusion of events (e.g., fun runs) or were given to participants to take with them to their activity for them to complete upon conclusion of this activity. All instructions were provided at the top of the questionnaire. The time elapsed from completion of their activity to completion of the questionnaire was recorded on the questionnaire. The average time was 24.6 minutes (SD = 25.2).

Statistical Analyses

Structural equation modeling procedures using maximum likelihood estimation with EQS 5.7b were employed. Missing data were managed via mean imputation. Items were loaded uniquely upon factors in all analyses. Factor variances were fixed to unity in all measurement-model analyses in order to identify the model. In higher-order factor analyses, an indicator on each factor was fixed to 1.0 along with the variance of the higher order factor to accomplish this same end.

An optimal set of indicators was selected from among existing and potential measurement items through an iterative process. A single item was introduced into a 36-item measurement model consistent with those previously reported (e.g., Jackson & Marsh, 1996; Marsh & Jackson, 1999) to allow the performance of the item to be evaluated (i.e., item loading, pattern of associated residuals, modification indices) within the context of all other construct indicators. This process was repeated until a conceptually and empirically optimal 36-item solution (4 items per factors) was identified for the 9-factor scales. In the few instances when item selection was statistically ambiguous, conceptual issues and the advantage of having a consistent set of indicators across inventory formats were the deciding factors.

We report the χ^2 test statistic and rely on the non-normed fit index (NNFI), the comparative fit index (CFI), and the root mean square error of approximation (RMSEA) to evaluate goodness of fit in these analyses (Hoyle & Panter, 1995). The χ^2, an absolute fit index, provides the basis for statistical tests of the lack of fit resulting from overidentifying restrictions placed on models. The NNFI estimates the relative improvement per degree of freedom of the target model over a baseline model. The CFI assesses the relative reduction in lack of fit as estimated by referencing the noncentral χ^2 of a target model to a baseline model. The RMSEA assesses the fit function of the target model adjusted by the degrees of freedom.

NNFI and CFI values exceeding .90 and .95 are typically taken to reflect acceptable and excellent fits to the data (Hoyle & Panter, 1995; Hu & Bentler, 1999). For the RMSEA, values of less than .05 and .08 are taken to reflect, respectively, a close fit and a reasonable fit (Browne & Cudeck, 1993). The RMSEA 90% confidence intervals are also provided to assist in interpreting these point estimates. Finally, evaluation of parameter estimates (i.e., factor loadings), modification indices, and the pattern of standardized residuals were also crucial in making decisions about the utility and statistical appropriateness of potential new items. Items were considered to be stronger indicators of their factor if they had larger factor loadings, modification indices suggesting that the item loaded simply, and residuals indicating a small discrepancy between observed correlations and model-reproduced correlations for the variable.

Results

Based on the item selection process described above, 5 of 13 new items were selected in Study 1 analyses to replace existing items in the measurement of the flow-experience scales. The new versions of the scales were given the titles FSS–2 and DFS–2. One new item replaced the problematic item, identified earlier, from the loss-of-self-consciousness scale. This new item was, "I am not concerned with how others may be evaluating me" (DFS wording). A second new item, "It is really clear to me how my performance is going" (FSS wording), replaced the problematic unambiguous-feedback item identified in the Introduction. Two new time-transformation items, "It feels like time goes by quickly," "I lose my normal awareness of time" (DFS wordings) replaced original items that focused on time slowing. Finally, a new sense-of-control item, "I have a sense of control over what I am doing" (DFS wording), replaced an original item that had focused on the notion of "total control."

Goodness-of-fit values for the final set of 36 items (31 original, 5 new) identified in the item evaluation analyses for the scales, for both the first-order factor

Table 1 Global Fit Indices for FSS–2 and DFS–2 Item Identification and Cross-Validation Analyses

		n	$\chi 2$	*df*	NNFI	CFI	RMSEA	90% CI
Measurement Model (9 first-order factors)								
Item ID	FSS–2	391	1171.026	558	.915	.925	.053	.049 – .057
Cross-val.	FSS–2	422	1177.558	558	.931	.939	.051	.047 – .055
Item ID	DFS–2	386	956.859	558	.943	.950	.043	.038 – .048
Cross-val.	DFS–2	574	1427.219	558	.901	.912	.052	.049 – .055
Higher-Order Factor Model (9 first-order factors, 1 second-order factor)								
Item ID	FSS–2	391	1266.189	585	.910	.917	.055	.050 – .059
Cross-val.	FSS–2	422	1305.374	585	.923	.929	.054	.050 – .058
Item ID	DFS–2	386	1063.348	585	.935	.940	.046	.042 – .050
Cross-val.	DFS–2	574	1606.487	585	.889	.897	.055	.052 – .058

Note: Item ID = Item identification model. Cross-val. = Cross-validation model.

model and higher order model, are listed in Table 1 in the appropriate Item Identification rows. Both the first order and higher order models demonstrated a good fit with NNFIs and CFIs all well above .9. RMSEA confidence intervals suggest reasonable (i.e., RMSEA <.08) or close (RMSEA <.05) fit for both models to FSS–2 and DFS–2 data in these analyses. Overall, the fit values were slightly better for the first-order factor model, but the difference is largely inconsequential.

Parameter estimates are listed for the higher order model in Table 2 in the Item Identification columns. The loadings of items on first-rder factors are all substantial, ranging from .51 to .89 for the FSS–2 (mean = .78). The corresponding DFS–2 loadings ranged from .59 to .86 (mean = .77). Correlations among the revised FSS–2 and DFS–2 first-order latent factors are listed in Table 3. They ranged from .13 to .76 (median r = .48) for the FSS–2 and from .24 to .78 (median r = .51) for the DFS–2. These values indicate that the nine flow factors, while sharing common variance as expected, measure reasonably unique constructs. Along with the item loadings on first-order factors in Table 2 are presented the loading of the first-order factors on the global flow factor. The first-order factor loadings on the flow latent factor range from .23 to .94 (mean = .66) for the FSS–2 and from .44 to .90 (mean = .71) for the DFS–2.

In summary, these results indicate that the DFS–2 and the FSS–2 demonstrate acceptable factorial validity for assessing dispositional and state flow, respectively. While encouraging, it is important to cross-validate the FSS–2 and DFS–2 models to ensure that the results observed in Study 1 are not sample-specific. Data for Study 1 were collected with 49 item versions of the scale. Cross-validation with the final 36-item versions of these scales is also important so as to ensure that items behave appropriately in the context of the final measurement presentation format. Study 2 was conducted to address these issues.

Table 2 Loadings From Item Identification and Cross-Validation Analyses of the FSS–2 and DFS–2

Factor	Item	FSS–2 Analyses Item ID	FSS–2 Analyses Cross-val.	DFS–2 Analyses Item ID	DFS–2 Analyses Cross-val.
F1 - Balance	FSS01	.574	.605	.622	.514
F1 - Balance	FSS10	.813	.812	.797	.707
F1 - Balance	FSS19	.808	.809	.852	.745
F1 - Balance	FSS28	.781	.763	.806	.771
F2 - Merging	FSS02	.629	.743	.666	.711
F2 - Merging	FSS11	.682	.848	.760	.733
F2 - Merging	FSS20	.840	.845	.832	.828
F2 - Merging	FSS29	.849	.864	.775	.832
F3 - Goals	FSS03	.725	.779	.677	.719
F3 - Goals	FSS12	.774	.850	.777	.747
F3 - Goals	FSS21	.763	.795	.783	.773
F3 - Goals	FSS30	.771	.758	.815	.709
F4 - Feedback	FSS04	.733	.736	.860	.728
F4 - Feedback	FSS13	.851	.785	.815	.797
F4 - Feedback	FSS22	.801	.853	.855	.824
F4 - Feedback	FSS31	.810	.832	.810	.788
F5 - Concentration	FSS05	.659	.775	.665	.611
F5 - Concentration	FSS14	.672	.697	.684	.643
F5 - Concentration	FSS23	.887	.866	.815	.806
F5 - Concentration	FSS32	.866	.892	.844	.780
F6 - Control	FSS06	.772	.799	.704	.675
F6 - Control	FSS15	.786	.799	.744	.718
F6 - Control	FSS24	.820	.842	.815	.771
F6 - Control	FSS33	.805	.786	.837	.718
F7 - Consciousness	FSS07	.822	.854	.742	.760
F7 - Consciousness	FSS16	.826	.912	.853	.812
F7 - Consciousness	FSS25	.786	.780	.780	.638
F7 - Consciousness	FSS34	.874	.903	.846	.823
F8 - Time	FSS08	.796	.813	.721	.733
F8 - Time	FSS17	.754	.871	.793	.826
F8 - Time	FSS26	.510	.433	.587	.606
F8 - Time	FSS35	.763	.722	.741	.732
F9 - Autotelic	FSS09	.755	.849	.702	.683
F9 - Autotelic	FSS18	.771	.771	.736	.550
F9 - Autotelic	FSS27	.835	.885	.779	.789
F9 - Autotelic	FSS36	.810	.898	.831	.779
F10 -Flow	F1	.779	.819	.846	.821
F10 -Flow	F2	.807	.704	.800	.718

(continued)

Table 2 (*Cont.*)

Factor	Item	FSS–2 Analyses Item ID	FSS–2 Analyses Cross-val.	DFS–2 Analyses Item ID	DFS–2 Analyses Cross-val.
F10 -Flow	F3	.736	.739	.806	.768
F10 -Flow	F4	.602	.597	.673	.707
F10 -Flow	F5	.763	.669	.776	.725
F10 -Flow	F6	.938	.895	.900	.908
F10 -Flow	F7	.446	.471	.441	.432
F10 -Flow	F8	.232	.208	.449	.300
F10 -Flow	F9	.660	.649	.706	.605

Table 3 Correlations Among FSS–2 and DFS–2 First-Order Latent Factors for Item Identification and Cross-Validation Analyses

	F1	F2	F3	F4	F5	F6	F7	F8	F9
FSS–2									
F1	1.000	.642	.559	.402	.491	.737	.374	.206.	.614
F2	.657	1.000	.538	.394	.371	.656	.348	.186	.358
F3	.584	.608	1.000	.598	.541	.620	.372	.155	.459
F4	.408	.509	.596	1.000	.488	.522	.312	.125	.285
F5	.510	.588	.574	.482	1.000	.630	.239	.063	.487
F6	.762	.760	.645	.528	.745	1.000	.425	.128	.579
F7	.235	.325	.316	.260	.397	.449	1.000	.078	.322
F8	.166	.252	.172	.163	.222	.140	.132	1.000	.268
F9	.584	.493	.475	.366	.502	.608	.336	.303	1.000
DFS–2									
F1	1.000	.707	.606	.620	.510	.727	.319	.248	.484
F2	.725	1.000	.505	.496	.380	.675	.327	.318	.373
F3	.642	.594	1.000	.636	.624	.654	.292	.167	.514
F4	.572	.516	.650	1.000	.520	.612	.249	.159	.333
F5	.594	.570	.670	.483	1.000	.715	.243	.208	.540
F6	.768	.776	.720	.623	.727	1.000	.480	.224	.540
F7	.348	.396	.269	.275	.398	.439	1.000	.196	.254
F8	.436	.309	.350	.237	.361	.376	.243	1.000	.345
F9	.648	.511	.638	.442	.611	.567	.251	.458	1.000

Note: Intercorrelations from item identification analyses are below diagonals while intercorrelations from cross-validation analyses are above diagonals.

STUDY 2

Method

Participants

A total of 897 respondents contributed data to the analyses of this investigation, 449 of whom provided FSS–2 data and 584 who contributed DFS–2 data. After deletion of inventories with systematic missing data (e.g., failure to complete a page of the inventory), 422 respondents provided FSS–2 data and 574 contributed DFS–2 data (99 provided both FSS–2 and DFS–2 data). Ages ranged from 16 to 82 years (M = 26.3, SD = 11.1, ~3% nonresponse rate). The gender breakdown was 48% males and 52% females (~3% gender nonresponse rate).

To be eligible for participation in Study 2, respondents again had to take part in physical activity at least twice a week. Participants engaged in a wide variety of physical activities. There were 27 activity types involving more than a single respondent (activities with only one respondent were collated into an "other" category). Once again, the types of activities ranged from highly competitive sports such as U.S. college football to exercise activities such as aerobics. This sample also included many dance (e.g., ballet, contemporary dance) and yoga participants. The 10 most common activities reported were running (n = 255), dance (n = 177), yoga (n = 99), triathlon (n = 56), Australian rules football (n = 51), basketball (n = 47), American football (n = 46), rugby (n = 33), track & field (n = 33), and soccer (n = 31). As with the first sample, there was also a wide range in levels of participation, including (but not limited to) international (5%), national (11%), U.S. college (16%), state (17%), and club or school (23%) involvement.

Instruments

Dispositional Measures. The first questionnaire included background demographic information and Dispositional Flow Scale–2 (DFS–2), which contained the 36 items identified in the analyses described in Study 1. As with the original DFS, respondents were instructed to think about how often they experienced each characteristic and to rate their responses on a 5-point Likert scale ranging from 1 = "never" to 5 = "always." Reliability estimates for the DFS–2 in Study 1 ranged from .81 to .90, with a mean α of .85.

State Measures. The second questionnaire contained questions about the activity just completed and Flow State Scale–2 (FSS–2). The activity questions were designed to ground the respondent in the experience of the event to which he or she was responding, with questions about the date, time of activity, time of questionnaire completion, and type of activity, as well as an open-ended description of his or her experience in the event. In addition, there were questions about how similar or different this event was from other times when the respondent took part in the activity, and he or she was asked to rate the challenges, skills, and his or her level of performance in the activity.

The FSS–2 contained the best 36 items (state version of items in DFS–2) from the analyses described in Study 1. Following the format of the FSS, respondents were asked to indicate the extent of their agreement with the items as characterizing their experience in the event just completed. Responses were on a 5-point Likert scale ranging from 1 = "strongly disagree" to 5 = strongly agree." Reliability estimates for the FSS–2 in Study 1 ranged from .80 to .90, with a mean α of .85.

Procedures

Participants were recruited from a variety of physical activity settings as well as from university Human Movement and Psychology classes. Standard introduction and information instructions were given to all participants, and informed consent procedures were adhered to. The dispositional measures were completed at a time separate from participants' participation in physical activity. A trained research assistant provided general instructions for completing the questionnaire. Specific instructions were included at the top of the questionnaire.

The FSS–2 was given to respondents to complete after participating in their main activity. Questionnaires were either distributed at the conclusion of events (e.g., a fun run) or were given to participants to take with them to their activity for them to complete upon conclusion of this activity. All instructions were provided at the top of the questionnaire. The time elapsed from completion of their activity to completion of the questionnaire was recorded on the questionnaire. The average time was 24.8 minutes ($SD = 26.1$).

Statistical Analyses

Structural equation modeling procedures using maximum likelihood estimation with EQS 5.7b were also employed in Study 2. Analyses and fit indices were all consistent with procedures described in Study 1, with a single exception. Analyses in Study 2 were conducted using means and covariances obtained via Graham and Hofer's (1995) EMCOV23 program. This program uses the EM algorithm (Dempster, Laird, & Rubin, 1977) which implements, by repeated imputation-estimation cycles, the full information maximum likelihood (FIML) approach for estimating means and covariance matrices from incomplete data. FIML treatment of missing data is a theory-based approach considered to be superior to the mean-imputation method (Wothke, 2000) employed in Study 1.

Results

Goodness of fit was evaluated for the DFS–2 and FSS–2 in cross-validation for both a 9-factor first-order factor model and a higher order model with a global flow factor. The results of these analyses are listed in Table 1 in the Cross-validation rows. The observed fit values for both the first-order and higher order models in these cross-validations are satisfactory. The DFS–2 and FSS–2 measurement models for the nine first-order factors exhibit NNFI and CFI values all exceeding .9. RMSEA point-estimate values for these models marginally exceed .05. Nonetheless, RMSEA 90% confidence intervals surrounding the point estimates suggest it would be intemperate to conclude that the RMSEA values do not indicate a close fit of the models to the data. The higher-order factor models exhibit NNFI and CFI values approximating or exceeding .9. The RMSEA point-estimate values for these models marginally exceed .05, and RMSEA 90% confidence intervals indicate that the models provide a reasonable if not close fit for the data. Overall, the fit values suggest a slightly better fit for the first-order factor models, particularly for the DFS–2.

Parameter estimates listed in Table 2 in the Cross-validation columns show good support for the nine flow dimensions. The loadings of items on the first-order factors are all substantial, ranging from .43 to .91 for the FSS–2 (mean = .80). The corresponding DFS–2 loadings ranged from .51 to .83 (mean = .73). Correlations among the first-order factors ranged from .06 to .74 (median r = .40) for the FSS–2 and from .16 to .73 (median r = .48) for the DFS–2 (see Table 3). The magnitude of these relationships again indicates that the flow subscales tap into reasonably unique aspects of the flow experience. Table 2 reveals that the loadings of the first-order factors on the global flow factor range from .21 to .90 (mean = .64) for the FSS–2 and from .30 to .91 (mean = .67) for the DFS–2.

Reliability estimates obtained for the FSS–2 in the cross-validation sample ranged from .80 to .92, with a mean α of .87. For the DFS–2, reliability estimates ranged from .78 to .86, with a mean α of .82.

Discussion

Confirmatory factor analyses from the two studies described in this report demonstrate that the DFS–2 and FSS–2 provide satisfactory and useful tools for assessing the constructs they were designed to measure: dispositional and state flow, respectively. The revised scales included items developed to enhance the conceptual coherence of the flow scales. Confirmatory factor analyses of data collected employing these items produced fit values that were usually better than, but at worst similar to, those obtained in research that employed the original versions of the scales (e.g., Jackson & Marsh, 1996; Marsh & Jackson, 1999). Not surprisingly, the best fit values were observed in item identification analyses, and these values were for the most part tangibly better than values observed in the past. More satisfying, however, was that the revised scales also performed well in cross-validation analyses.

These findings are encouraging and it is reasonable to conclude that the revised versions of Jackson's (Jackson & Marsh, 1996; Jackson et al., 1998; Marsh & Jackson, 1999) flow scales show overall improvement at conceptual and statistical levels. These scales should facilitate the self-report assessment of the nine flow dimensions and facilitate research into optimal experience in physical activity settings, although certainly the utility of the new flow scales is an empirical question for future research to address.

The item identification analyses did not reveal any substantial weaknesses statistically with the scales, although the higher-order factor loadings of time transformation for the FSS–2 still remain relatively weak. At the item level, one time-transformation item had a relatively weak factor loading for the FSS–2 cross-validation analysis. This was one of the new items that focused on time passing quickly. Since the loading on the DFS–2 cross-validation analysis for this item remained reasonable, it is unclear whether the item is problematic or simply dependent on situational variation that is part of FSS sampling and the flow experience in general.

As with the original scales, the recommended procedure for the DFS–2 and FSS–2 is to use factor level scores rather than a single global score. The purpose of developing a multidimensional scale to assess flow was to facilitate assessment of the construct at the level of the nine flow dimensions. This approach to evaluation of experiences in physical activity settings can provide more information about the nature of the flow experience than would be available in a single score approach. Nonetheless, it is recognized that global flow assessment is useful in addressing some research questions.

The results of analyses conducted in the present study, along with research conducted with the original scales, offer some support for the use of a global flow score. The higher-order factor loadings do indicate, however, that nine flow dimensions contribute unequally to global flow factor. As discussed by Jackson and Marsh (1996), it may be more accurate to allow the weighting of specific components to vary appropriately in different applications. Jackson and Marsh (1996) and Marsh and Jackson (1999) discuss in detail some issues related to the use of the 9-factor versus the global factor approach. Those observations are relevant to the use of the DFS–2 and FSS–2.

It is interesting to speculate about the reasons for the differential loadings of the nine flow factors on the higher-order flow factor. The largely consistent pattern of loadings is evident in research with the original versions of the scales and with those observed with the DFS–2 and FSS–2 in this investigation. Loss of self-consciousness and time transformation consistently tend to exhibit low loadings.

Modest but substantial improvement was observed in the DFS–2 time-transformation loading with the addition of new items to this scale. As discussed by Jackson and colleagues (e.g., Jackson & Marsh, 1996; Jackson et al., 1998), it may be that certain flow dimensions are more relevant to physical activity flow experiences than others. For example, perhaps a certain degree of self-consciousness may be a necessary part of the sport flow experience. At least some minimal awareness of body and self is necessary for competent performance in many of these physical activities. Part of the challenge of these activities is to be aware and able to evaluate how others are viewing one's performance without internalizing or overly focusing on such information.

Awareness of time may also be a relevant consideration to athletic situations where flow occurs. It is of course critical to some sports where the time clock is part of the structure of the game (e.g., basketball) or is part of the performance evaluation (e.g., swimming race). In some sports, however, time does not play such a critical role (e.g, rock climbing), and it may be that this dimension of time transformation is endorsed differently depending on the type of physical activity.

There are other possible explanations for the pattern of associations that these subscales exhibit with the global flow construct. It may be that the items used to assess these two dimensions are not relevant for certain activities or perhaps were not fully understood by the participants. For example, a figure skater has to be concerned with how she presents herself during her performance, so the item "I am not concerned with how I am presenting myself" may not be a relevant indicator of a flow-like loss of self-consciousness during her performance. In terms of complexity of constructs, the loss-of-self-consciousness and time-transformation dimensions represent more ephemeral themes than do dimensions such as clear goals and total concentration on the task at hand, the latter of which load up substantially on the higher-order flow factor. These matters certainly merit evaluation in future research efforts.

It is important to note that the newly revised loss-of-self-consciousness scale has a self-presentational flavor. This is perhaps a narrower conceptualization of loss of self-consciousness than was implied by Csikszentmihalyi (1990). He has referred to a lack of focus upon information we normally use to represent to ourselves who we are when experiencing a loss of self-consciousness. The items that represent loss of self-consciousness in the DFS–2 and FSS–2 focus primarily on the loss of concern with evaluation of self by others. Perhaps this is a central consideration when evaluating self, particularly in the public realm of sport and physical activity. Nonetheless, the similarity in items for this factor means that only a limited assessment of loss of self-consciousness is provided.

Two important strengths of the evaluations of the FSS–2 and DFS–2 in this report should be mentioned. First, despite evidence suggesting that the original versions of these measures were satisfactory, analyses were cross-validated with data collected from a second independent sample to ensure that item identification results were not sample-specific or confounded by the inclusion of excess items. Second, the nature of the samples themselves was considerably broader than studies with the original scales. A large number of participants were involved who participated in a diverse range of activities. In addition to sport participants, the samples included exercise and dance participants. This diversity in sampling indicates that the scales are likely applicable across a wide array of activities.

The research described herein is of course subject to limitations, and many are similar to those reported in earlier flow scale investigations (Jackson & Marsh, 1996; Jackson et al., 1998; 2001). In this study we sought to collect FSS–2 data in a way that minimized the extent of retrospective recall required. Although we tried to have people complete the FSS–2 inventory promptly after participating an activity, we had only limited control over when the state responses were actually completed. A further issue discussed by Jackson and colleagues relates to how well flow can be assessed by a quantitative self-report format. As has been raised in other discussions, the scales represent an attempt to quantify experiential states, and this is not an easy task.

We agree with Csikszentmihalyi (1992) that one should not place too much weight on any empirical measure of flow, so as not to lessen or lose the experience by reducing it to scores on questionnaires. However, we believe that through measures such as those described in this report we can open up possibilities for investigating flow in creative ways and in diverse settings. We present the flow scales as means of tapping into the expression of flow in physical activity. Just how well they can do this is up to future research to address. The scales are presented here as the next step in the development of a self-report approach to flow assessment in physical activity.

There are many directions future research could take to shed more light on the flow experience in physical activity. For example, as mentioned, unraveling the relative relevance of the nine flow dimensions to different individuals in different activity settings would be an important area to address. Relating the flow dimensions to relevant criteria would add to the external validity of the scales and increase knowledge of relevant antecedents to, and consequences of, the flow experience. Other possible paths to follow include comparisons of groups predicted to differ in flow experience, as well as using DFS–2 and FSS–2 responses as outcome variables in interventions designed to increase flow. Approaching the study of flow creatively and from a multidisciplinary perspective offers the greatest promise of developing a better understanding of the experience of this optimal psychological state in physical activity.

References

Begly, G. (1979). *A self-report measure to assess flow in physical activities*. Unpublished master's thesis. The Pennsylvania State University, University Park.

Browne, M.W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models*. Newbury Park, CA: Sage.

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety*. San Francisco: Jossey Bass.

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., & Csikszentmihalyi, I. (Eds.) (1988). *Optimal experience: Psychological studies of flow in consciousness*. Cambridge: Cambridge University Press.

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the Experience Sampling Method. *Journal of Nervous and Mental Disease*, **175**, 526-536.

Dempster, A.P., Laird, N.M. & Rubin, D.B. (1977) Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39**, 1-38.

Graham, J.W., & Hofer, S.M. (1995). *Covariance estimation with missing values: An application of the EM algorithm* [Online]. Available: ftp://ftp.cac.psu.edu/pub/people/jwg4/dos/emcov.txt

Hoyle, R.H., & Panter, A.T. (1995). Writing about structural equation models. In R.H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 158-176). London: Sage.

Hu, L., & Bentler, P.M. (1999). Cutoff 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. (1995). Factors influencing the occurrence of flow in elite athletes. *Journal of Applied Sport Psychology*, **7**, 138-166.

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., Kimiecik, J.C., Ford, S., & Marsh, H.W. (1998). Psychological correlates 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**, 154-178.

Logan, R. (1988). Flow in solitary ordeals. In M. Csikszentmihalyi & I. Csikszentmihalyi (Eds.), *Optimal experience: Psychological studies of flow in consciousness* (pp. 172-180). Cambridge: Cambridge University Press.

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

Marsh, H.W. (1998). Foreword. In J.L. Duda (Ed.), *Advances in sport and exercise psychology measurement*. Morgantown, WV: Fitness Information Technology.

Marsh, H.W. (2000, Aug. 30). *Conventional fit criteria too restrictive* [Formal discussion initiation to Structural Equation Modeling Discussion Group]. Retrieved Aug. 30, 2000, from the World Wide Web. URL: http://bama.ua.edu/archives/semnet.html

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

Privette, G., (1984). *Experience questionnaire*. Pensacola: University of West Florida.

Privette, G., & Bundrick, C.M. (1991). Peak experience, peak performance, and flow. *Journal of Social Behavior and Personality*, **6**, 169-188.

Wothke, W. (2000). Longitudinal and multigroup modeling with missing data. In T.D. Little, K. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples*. Mahwah, NJ: Erlbaum.

Vlachopoulos, S.P., Karageorghis, C.I., & Terry, P.C. (2000). Hierarchical confirmatory factor analysis of the Flow State Scale in exercise. *Journal of Sports Sciences*, **18**, 815-823.

Acknowledgments

The authors would like to acknowledge the research assistance of Bill Wrigley and Ian Thomas, who helped with various phases of this investigation. The authors would also like to acknowledge Lew Curry and Peter Haberl for providing U.S. sport data, and Herb Marsh for feedback on statistical issues.

Manuscript submitted: July 30, 2001
Revision accepted: January 16, 2002