Construct Validity Evidence for a Measure of Peer Athletic Reputation (PAthR) in High School Physical Education

Citation: Howle, T. C., Whipp, P. R., Gairns, F., Thornton, A. L., Rebar, A. L., & Jackson, B. (2016). Construct validity evidence for a measure of peer athletic reputation (PAthR) in high school physical education. Psychology of Sport and Exercise, 24, 9-17
Abstract

Grounded in the work that has been conducted on peer academic reputation in educational settings, our aim was to develop and provide initial construct validity evidence for a modified instrument designed to assess peer athletic reputation (PAthR). In phase 1, we devised a series of items to assess PAthR. In phase 2 \((n = 399)\), and approximately one month later in phase 3 \((n = 374)\), female high school physical education (PE) students completed a series of questionnaires that included the PAthR instrument. Analyses indicated that PAthR was (a) represented as a unitary construct, (b) positively associated across time points, (c) positively associated with indicators of desirable peer interactions (e.g., peer relatedness support, peer relatedness need satisfaction) and PE experiences (e.g., intrinsic motivation, perceived competence), and (d) negatively associated with indices considered to be less adaptive in nature (i.e., anxiety, amotivation). These findings provide preliminary support for a novel approach to the measurement of peer reputation in PE, and demonstrate evidence of the nomological net with which this concept may be associated.

Keywords: Competence; Ability; Perceptions
**Construct Validity Evidence for a Measure of Peer Athletic Reputation (PAthR) in High School Physical Education**

The contribution of teachers, coaches, and parents in supporting physical activity experiences during childhood and adolescence is well established (e.g., Ornelas, Perreira, & Ayala, 2007; Saelens & Kerr, 2008; Sallis, Prochaska, & Taylor, 2000). In recent years, however, researchers have begun to more closely examine the unique role played by peers in this respect (see Smith & McDonough, 2008). In this manuscript, we seek to add to this literature by exploring the nature, measurement, and correlates of peer athletic reputation. There has been increasing recognition among educational psychologists over the last decade regarding the developmental significance of peer academic reputations (PAR; e.g., Gest, Domitrovich, & Welsh, 2005; Gest, Rulison, Davidson, & Welsh, 2008); that is, the aggregated judgments of peers regarding a focal individual’s academic competence. Gest and colleagues (Gest et al., 2008), for example, noted that individuals might base their interaction decisions and behavior (e.g., regarding cooperation, communication, attention, support, praise/derogation) on the reputational perceptions that they (and/or others) hold about specific peers’ academic competence. Accordingly, by shaping interpersonal relations between peers, it has been proposed that this reputational variable might predict changes in self-perceptions and behavior (e.g., self-esteem, self-concept, effort, performance), in much the same way that one’s social reputation has been shown to influence social adaptation (e.g., Hymel, Wagner, & Butler, 1990).

There is empirical support for these assertions; it has been demonstrated, for instance, that PAR may predict outcomes such as an individual’s self-concept, engagement, skill development, and performance (e.g., Chen, Hughes, Liew, & Kwok, 2010; Gest et al., 2005, 2008; Hughes, Dyer, Luo, & Kwok, 2009). To date, however, researchers are yet to examine the extent to which peer-group assessments are made in relation to individuals’ athletic
Peer athletic reputation (hereafter, PAthR) represents a useful marker of an individual’s athletic ability, and that PAthR may operate through comparable processes as PAR in shaping physical activity-related outcomes (e.g., engagement, motivation, skill development). Guided by previous work in classroom-based contexts, we therefore defined PAthR as the aggregate judgment of one’s peers – in this instance, one’s physical education (PE) classmates – with respect to one’s athletic ability.

Peer reputations, whether academic or athletic in nature, emerge through the access that peers have to information regarding the ability of group members or classmates (Gest et al., 2005). In classroom settings, the time that peers spend working alongside one another allows for the detection and social comparison of ability (Ruble, Boggiano, Feldman, & Loebl, 1980), and evidence suggests that adolescents in physical activity settings also engage in these detection and comparison processes (cf., Chanal, Marsh, Sarrazin, & Bois, 2005). Importantly, not only do peers make social comparisons, but the assessments that they make about group members may reflect unique insight into individuals’ ability. Indeed, research findings show that although peer-based ratings of ability are positively correlated with teacher ratings, peer ratings contribute independently to the prediction of outcomes such as self-concept and performance (Gest et al., 2005, 2008).

The outcomes associated with these peer-group inferences are theorized to arise out of reputation-dependent differences in the way in which individuals respond to, and interact with, their peers. Although group interaction processes have not been considered through the lens of PAthR before, researchers have demonstrated that the characteristics of one’s peer environment may drive personal outcomes. Research involving adolescent participants, for example, indicates that positive peer-based social-comparative feedback may enhance motor learning (Ávila, Chiviacowsky, Wulf, & Lewthwaite, 2012), and that the presence of a task-involving climate created by peers may positively influence intrinsic motivation for physical
activity (Jõesaar, Hein, & Hagger, 2012). Balanced alongside this research is evidence that a task-involving peer-created PE climate positively predicts physical activity behaviour but not physical self-concept (Taylor, Spray, & Pearson, 2014). Accounting for other peer-related factors, such as PAthR, may provide additional insight into the relationship between the peer-created climate and individuals’ physical activity outcomes.

It is also possible that, on the basis of shaping others’ interaction styles and behavior, PAthR might be important in helping individuals interpret how confident their peers are in their ability and what their peers think about them. Self-efficacy theorists contend that individuals’ inferences about how confident their peers – as a whole – are in their ability (i.e., relation-inferred self-efficacy, or RISE) may positively predict their own self-efficacy beliefs (Lent & Lopez, 2002). Drawing from social cognitive theory (Bandura, 1997), these confidence beliefs may then positively influence one’s effort, persistence, and performance in that domain. Jackson and colleagues (Jackson, Gucciardi, Lonsdale, Whipp, & Dimmock, 2014) applied Lent and Lopez’s (2002) relational efficacy framework to study the nature of peer-focused RISE perceptions within the physical activity domain. Their research supported a model in which positive predictive pathways existed from PE students’ peer-focused RISE appraisals to their PE self-efficacy beliefs, and subsequently to a range of positive outcomes including enjoyment, persistence intentions, attitudes, and achievement. Reflected appraisals (i.e., individuals’ perceptions of what others think about them) may be an alternate mechanism through which PAthR influences self-perceptions. Athletes’ teammate-related reflected appraisals, for example, have been found to positively predict athletes’ own perceptions of their sport self-competence (Trouilloud & Amiel, 2011), supporting the notion that individuals may develop self-perceptions that are internalized from their beliefs about the way in which others perceive them (Felson, 1993). Researchers, however, have not considered whether an individual’s PAthR contributed to favorable (or unfavorable) peer-
focused RISE beliefs and reflected appraisals, or the extent to which reputation scores might
independently align with other processes, as has been demonstrated in classroom (i.e., non-
PE) settings.

There is sufficient empirical evidence to suggest that the study of PAthR may hold
merit in understanding individuals’ physical activity experiences (Gest et al., 2005, 2008). To
date, however, PAthR is yet to be operationalized in physical activity settings, and the
development of such work is limited by the absence of a valid measure of the construct. In
the present manuscript, we seek to address this limitation by developing and examining
preliminary construct validity evidence for an instrument designed to assess PAthR. Our
focus on instrument development and validity evidence was guided by Messick’s (1995)
unified view of construct validity. Messick contended that researchers interested in construct
validation should account for different aspects of validity during instrument development and
testing. Construct validity is considered to be represented by content (i.e., item relevance,
representativeness, and technical quality), substantive (i.e., the theoretical rationale),
structural (i.e., the fidelity of the scoring structure), generalizability (i.e., that score properties
and interpretations generalize across population groups, settings, tasks, and/or time), external
(i.e., convergent and discriminant validity evidence), and consequential (i.e., the value of
score interpretation and the consequences of instrument use) aspects. Measure development is
an iterative process (Messick, 1995), meaning that validation efforts should be ongoing so as
to examine these different aspects over time. This has important implications for the present
study insofar as our aim was to take an initial step in the validation process by developing a
PAthR measure and examining evidence for the content, substantive, structural,
generalizability, and external aspects of construct validity. We aimed to achieve these goals
by structuring our investigation into three phases.
In phase 1, we devised a pool of items to assess PAthR and endeavoured to provide evidence for the content and substantive aspects of validity for scores derived from this measure. In phase 2, we administered the PAthR measure to a sample of high school PE students and examined evidence for the structural and external aspects of validity. The choice of sample (i.e., PE students) was made so as to be consistent with previous work that has demonstrated the significance of reputational inferences among this cohort (e.g., Chen et al., 2010; Gest et al., 2005, 2008; Hughes et al., 2009). The structural aspect of validity was investigated through the use of confirmatory factor analysis (CFA). We anticipated that we would observe support for a single factor (i.e., PAthR) model, reflecting Gest et al.’s (2008) view that peer-reputation can be considered as a unitary index. It is worth briefly noting that the proposed measure focused on individuals’ general athletic reputation, rather than considering separate aspects of athleticism (e.g., talent and effort). This approach was consistent with Gest et al.’s (2005, 2008) assessments of academic reputation. Additionally, our focus was on examining PAthR within an all-female sample. Relative to males, females may place a stronger emphasis on defining and evaluating themselves in terms of their interpersonal relationships (Rose & Rudolph, 2006; Rudolph & Conley, 2005). As such, PAthR may be more important to one’s self-concept for females and associations between PAthR and prospective outcome variables (e.g., peer group perceptions, PE experiences) may be more apparent within female samples.

On the basis of previous research (e.g., Jackson et al., 2014; Ntoumanis, 2012; Smith, 2003), we also sought to examine evidence for the external aspect of validity in phase 2 by exploring associations between PAthR and a range of variables that represent markers of peer interaction processes (i.e., peer-focused relatedness support, peer-focused relatedness need satisfaction, peer-focused RISE, peer-focused social anxiety) or PE/physical activity experiences (i.e., competence, intrinsic motivation, amotivation, PE engagement, leisure-time
physical activity; LTPA). The inclusion of peer-focused measures allowed for the
investigation of how peer reputation in PE aligns with (and is distinct from) other important
peer-focused PE perceptions. Meanwhile, the inclusion of general PE/activity-related
measures enabled us to consider the potential engagement-related outcomes with which
PATHR may align. Drawing from the work of Gest et al. (2005, 2008), we hypothesized that
positive associations would be observed between an individual’s PATHR score and variables
that were indicative of positive peer interactions (i.e., peer relatedness support, peer RISE,
peer relatedness need satisfaction) and PE/physical activity experiences (i.e., perceived
competence, intrinsic motivation, PE engagement, LTPA). Conversely, negative associations
were expected between an individual’s PATHR score and variables that reflected a less
positive peer/PE experience (i.e., social anxiety, amotivation).

In phase 3, we focused on addressing the generalizability aspect of validity by
examining the stability of PATHR scores and external validity associations across time. This
aim was addressed by resampling phase 2 participants at a later time point and with respect to
a different focal PE activity. Stability across time is an important part of the generalizability
aspect of validity (Messick, 1995). We expected a positive association between an
individual’s PATHR across time points, allowing for change in peer reputation that is thought
to occur over time and across activities, but also acknowledging that a relatively stable,
underlying level of athletic competence may be present for participants (Gest et al., 2005).
Expectations for external validity associations were the same as those outlined for phase 2.

**Phase 1: Item generation**

We generated four items (i.e., “is good at the sport you are currently doing”, “is not
good at the sport you are currently doing”, “does activities and skills well in the sport you are
currently doing”, and “does not do activities and drills well in the sport you are currently
doing”) to assess PATHR within the context of high school PE. In generating the four PATHR
items, we attended to the content and substantive aspects of validity by focusing on item technical quality, relevance, and representativeness. Judgements about item relevance and representativeness should be informed by existing evidence of the structure of the relevant behavioural domain and the associated test responses (Messick, 1994). In this respect, we were guided by Gest and colleagues’ (2005, 2008) previous work which supported the appropriateness of a peer-reputation measure that comprised four total items, with two items positively-worded and two items negatively-worded. We sought to generate items that were appropriate for the assessment of athletic reputation rather than academic reputation by modifying items used by Gest et al. (2005, 2008) for use in the PE context. Item modification was performed by the authorship team following a period of internal discussion. The authors have experience with performing research in PE settings, instructing (e.g., coaching) adolescents as part of physical activity programs, and developing psychological measures for the PE context. To make item modifications, we drew from material (Jackson, Whipp, Chua, Pengelley, & Beauchamp, 2012) that had a similar conceptual (i.e., peer-related efficacy beliefs), contextual (i.e., high school PE), and theoretical (i.e., how peer views may shape self-concept and behaviour) focus to the present work. The modified and original items are presented in Table 1. We were also mindful of generating items that were easily understood by the target population. The four generated items had a 4.2 Flesch-Kincaid grade level, indicating that they were appropriate for use for students at and above U.S. grades 4 to 5.

**Phase 2: Testing structural and external aspects of validity**

Having developed items intended to assess PAthR, and considered the content and substantive aspects of validity in phase 1, our focus in phase 2 was on examining the structural and external aspects of validity for this measure. This represents the first time that the PAthR measure has been administered.
Participants and procedure. A total of 399 female participants ($M_{age} = 13.43, SD = 1.21$) were recruited from 19 separate compulsory PE classes within grade 7 ($n = 137$), 8 ($n = 90$), 9 ($n = 82$), and 10 ($n = 90$) at an Australian all girls’ independent school. This represented a 95% response rate from the 418 total students enrolled in these classes. Following institutional ethics approval, information sheets were sent to the school principal and an invitation to participate in the study was extended. Information sheets detailed the study purpose, design, and procedure. Having obtained consent from the principal, information sheets were then sent to prospective teachers, students, and parents/guardians similarly outlining the nature of the study. All PE teachers of grade 7 to 10 students received an invitation to participate, and classes that were taught by these teachers were selected at random. The researchers arranged a suitable time with teachers to visit the school, and two- to-three weeks prior to the visit, an information letter was sent home to parents with the invitation to allow them to withdraw their daughter from the investigation should they wish. During class time, students and teachers first provided their informed consent, before students completed the questionnaire measures outlined below. Prior to data collection, students were informed in writing and reminded verbally that they had the right to refuse participation, refuse to answer any question, and withdraw at any time. They were also informed that all information would remain confidential and that their decisions would not impact their standing with their teacher or school.

To avoid overburdening students and disrupting teachers’ lesson plans, participants completed the questionnaire measures over two lessons (separated by two or three days) within a given week. In the first lesson, students provided demographic information and reported on the LTPA that they performed over the previous week. At the start of the second lesson, participants completed a battery of measures that included relatedness support and RISE measures, and at the end of the lesson, students completed the PAtHR, need satisfaction,
motivation, and anxiety measures. During this same lesson, teachers rated their students’ behavioral engagement in their PE classes over that week.

**Measures.**

**PPathR.** Consistent with previous peer reputation work in the academic domain (Gest et al., 2008), participants were provided with a tabulated single-page list of all individuals within their class and asked to respond to the four items that we had generated in phase 1 to assess PPathR (items were presented in the top row across four columns, with all student names appearing in repeat fashion in each column beneath the item). Participants were asked to indicate (by circling) the names of their classmates that they thought matched each item. The use of peer-based nomination methods has been reviewed extensively elsewhere and the appropriateness of this method supported (c.f., Terry, 2000). For the present study, instructions specified that participants could circle as many names as they liked (with the exception of their own name), and that a given student could be circled as many times as was deemed appropriate across the four items. For each item, an individual’s score could therefore range from 0 (i.e., no classmate nominated that student for that item) to the total number of students within the class minus one (i.e., the entire class nominated that student for that item, with the exception of the individual him/herself). The negatively worded items were reverse scored and a total mean score was calculated for each participant, whereby higher scores reflected a more positive reputation score. In addition, to account for varying class size, we followed the methodology used by Gest and colleagues and computed PPathR scores that were standardized within each class. The Cronbach’s alpha for the PPathR measure was .94.

**Peer-focused relatedness support.** Students’ perceptions of their peers’ relatedness supportive behaviour were assessed using Standage, Duda, and Ntoumanis’ (2005) measure. Items were modified to be contextually appropriate by including references to classmates
rather than teacher behaviour. Using a Likert-type scale with anchors at 1 (strongly disagree) and 7 (strongly agree), students responded to five items (e.g., “my classmates have respect for me”), following the stem, “At the moment, in my PE class…” Standage et al. found support for the external and structural aspects of validity for scores derived from this measure. In the present investigation, we observed an acceptable internal consistency estimate for the measure derived from this instrument (α = .93).

**Peer-focused RISE.** Peer-focused RISE beliefs (i.e., one’s evaluation of how confident one’s classmates as a whole are in one’s PE ability) were assessed using Jackson et al.’s (2012) nine-item instrument. We again included contextual modifications to orient participants to consider peer- rather than teacher-focused perceptions (see Jackson et al., 2014), and participants responded on a scale that ranged from 1 (no confidence at all) to 5 (complete confidence). Participants were requested to estimate, “right at this moment in time, how confident do you think your classmates as a whole are in your ability to…” (e.g., “perform all the skills you are taught in PE”). Internal consistency and construct validity evidence has previously been reported by Jackson et al. (2012), and in the present investigation we observed an acceptable level of internal consistency for the peer-focused RISE measure derived from this instrument (α = .88).

**Peer-focused relatedness need satisfaction.** Relatedness need satisfaction regarding one’s peers was assessed with Richer and Vallerand’s (1998) five-item instrument. Using the stem “At the moment, with the other students in this PE class, I feel…”, items included “listened to” and “valued”. The response scale was anchored at 1 (strongly disagree) and 7 (strongly agree). Support for the construct validity of scores derived from Richer and Vallerand’s (1998) instrument has been presented in previous investigations (e.g., Standage, Duda, & Ntoumanis, 2003). We observed acceptable internal consistency for the relatedness measure derived from this instrument (α = .95).
Peer-focused social anxiety. Peer-focused social anxiety (i.e., students’ concerns regarding their peers’ impressions of them during their PE classes) was measured using four items that were drawn from Martin and Fox’s (2001) instrument. Contextual modifications were made such that participants were oriented to consider their classmates (e.g., “I am concerned about looking uncoordinated in front of my classmates”). Participants indicated their responses on a 5-point Likert-type scale with anchors at 1 (not at all concerned) and 5 (extreme concern) following a stem that requested they think about their feelings at that moment in time in their PE lessons. Consistent with Martin and Fox’s work, an acceptable internal consistency estimate ($\alpha = .92$) was observed for the measure derived from this instrument in the present study.

Competence. We assessed students’ perceptions of competence in PE using the perceived competence subscale from the intrinsic motivation inventory (Ryan, 1982). This measure includes six items (e.g., “I think I am pretty good at this activity”), and to orient participants toward the activity they were currently studying (i.e., the same referent as the PAthR assessment), we used the stem, “right now, in relation to the activity you are currently doing in this PE class…” The response scale was anchored at 1 (not at all true) and 7 (very true). The reliability and validity of scores derived from this instrument have been demonstrated previously (e.g., Standage et al., 2005), and we observed an acceptable level of internal consistency for the competence measure derived from this instrument ($\alpha = .91$).

Intrinsic motivation and amotivation. Participants’ intrinsic motivation and amotivation for PE were assessed using two subscales from the Perceived Locus of Causality instrument (Goudas, Biddle, & Fox, 1994). Using the stem, “at the moment, I take part in PE…”’, participants responded to statements reflecting intrinsic motivation (4 items; e.g., “because I enjoy learning new skills”) and amotivation (4 items; e.g., “but I don’t really know why”) that each contain 4 items. The 7-point response scale was anchored at 1 (strongly
disagree) and 7 (strongly agree). Support for the psychometric properties of the instrument has previously been demonstrated (e.g., Lonsdale, Sabiston, Taylor, & Ntoumanis, 2011) and in the present study the internal consistency of measures derived from these instruments was supported ($\alpha = .84$ for amotivation and .91 for intrinsic motivation). We measured intrinsic motivation as it represents the highest ‘quality’ form of motivation outlined within self-determination theory, and assessed amotivation on the grounds that it is the least desirable type (i.e., an absence) of motivation (Deci & Ryan, 2000). We excluded other motivational regulations given that (a) our focus was not on testing relations with all types of motivation, and (b) we sought to minimize participant burden.

**Behavioral engagement.** In an attempt to avoid self-report bias, we invited teachers to rate the degree of students’ in-class behavioral engagement. At the request of teachers (to minimize burden), a single-item measure was used. Teachers were provided with a roll of the students in their class, and for each pupil they were asked, “Over this week, what level of engagement has this student shown in your PE class?” The response scale had anchors at 1 (no engagement), 4 (average engagement), and 7 (very high level of engagement), and teachers were asked to rate each student based on the intensity of their participation (relative to their classmates) within the last week. Criterion validity for similar single-item measures has been found in previous research (e.g., Ntoumanis, 2005).

**LTPA.** LTPA was assessed using the Leisure-Time Exercise Questionnaire (Godin & Shephard, 1985). Students were provided with definitions and examples of mild, moderate, and vigorous physical activity, and asked to indicate how many sessions (> 20 minutes) of those activity types they had performed over the previous week. Instructions specified that students should not include physical activity that was performed as part of their PE classes or was otherwise a compulsory school requirement. A single LTPA score was calculated using Godin and Shephard’s recommended formula (i.e., $9 \times$ number of vigorous bouts $+ 5 \times$...
number of moderate bouts + 3 x number of mild bouts). Researchers who have used this measure with a similar sample have indicated evidence in support of it as an appropriate and valid instrument (e.g., Hagger, Chatzisarantis, Barkoukis, Wang, & Baranowski, 2005).

**Data analysis.** Data analysis was performed using IBM SPSS Version 21, *Mplus* version 7.4, and *R* version 3.1.3 (R Core Team, 2015). We first screened data and examined item-level descriptive statistics and normality information (i.e., skew and kurtosis values). We conducted intraclass correlations (ICCs) to investigate how much variability at the item- and scale-level was accounted for by class membership.

We then investigated support for the structural aspect of validity by using a CFA model in which each of the P Ath R items were specified to load on a single factor (i.e., representing P Ath R). The residuals of the two reverse-scored items were specified to be correlated and item intercepts were fixed to zero (see Kline, 2011; Muthén & Muthén, 2012). As a supplementary analysis, we also performed a multilevel CFA (see supplementary material). Model fit was examined using the $\chi^2$ goodness-of-fit index, the comparative fit index (CFI), Tucker-Lewis index (TLI), and the standardized root mean square residual (SRMR). The root mean square error of approximation (RMSEA) was not considered because the model had only a few (i.e., five) degrees of freedom. RMSEA is not recommended for use when models only have few degrees of freedom given that it too often indicates poor fit when the model is correctly specified (see Kenny, Kaniskan, & McCoach, 2014). Guided by Marsh (2007), an excellent fitting model was judged to be one with CFI and TLI values ≥ .95, and a SRMR value ≤ .08 (Hu & Bentler, 1999; Marsh, Hau, & Wen, 2004; Tabachnick & Fidell, 2007). In interpreting the strength of factor loadings, we were guided by Comrey and Lee’s (1992) recommendations (i.e., >.55 = good, .45 - .55 = fair, .32 - .45 = poor, < .32 = should not be interpreted). We also examined internal reliability
information (i.e., Cronbach’s alpha) for scores derived from the PAthR measure. A value of ≥ .70 was judged to be acceptable (Nunnally & Bernstein, 1994).

To examine support for the external aspect of validity, we used hierarchical linear modelling that accounted for nesting at the class level (Bates, Maechler, & Bolker, 2011) with 95% confidence intervals estimated from 100 posterior simulations (Gelman & Hill, 2006; Gelman et al., 2012). Random effects models were set to test bivariate associations between the PAthR score and peer-focused relatedness support, peer-focused RISE, peer-focused relatedness and need satisfaction, peer-focused social anxiety, competence, amotivation, intrinsic motivation, PE engagement (teacher rating), and LTPA. The variable scales were left unadjusted and the intercept and slope were allowed to vary across classes. Missing data, which comprised 1.42% of total data, were dealt with using the expectation maximization (EM) procedure (see Graham, 2009). Five univariate outliers (i.e., extreme z scores) and 19 multivariate outliers (i.e., Mahalanobis distance at $p < .001$) were removed.

**Results**

Regarding PAthR items, skewness (i.e., range -.83 to .30) and kurtosis (i.e., range -.94 to .56) statistics were within the acceptable range and did not indicate any problematic normality violations (Tabachnick & Fidell, 2007). PAthR item means (prior to being standardized) ranged from 9.51 to 17.67 ($SD$ range 4.69 to 6.33). The ICCs indicated that 12.52% of variability in PAthR was accounted for at the class level. ICCs were similar across the four items (ICC’s: 10.52% to 19.94%). Corrected item-total correlations ranged from .82 to .89, and the deletion of PAthR items would not have improved the measure’s internal reliability. Descriptive information is presented in Table 1. Consistent with the existing PAR literature (Gest et al., 2008), this table includes non-standardized information so as to aid ease of interpretation. Results from the factor analysis revealed support for the single-factor structure of PAthR. With the exception of the chi-square value, the fit indices showed that the
model was an excellent fit to the data, $\chi^2 (5) = 18.05, p < .003, CFI = .990, TLI = .988, SRMR = .009$. Standardized factor loadings (see Table 1) were considered good (Comrey & Lee, 1992) and ranged from .81 to .997.

Bivariate associations between PAthR and other variables – after accounting for nesting at the class level – are presented in Table 2. The results indicated that PAthR was positively associated with markers of positive peer interactions (i.e., relatedness support, RISE, relatedness need satisfaction), and indices relating to PE/physical activity experiences (i.e., intrinsic motivation, PE engagement, LTPA). Students’ athletic reputation scores were also negatively associated with undesirable peer- or PE-related outcomes (i.e., social anxiety, amotivation). Taken together, these associations indicated that when a student was rated as having a more positive athletic reputation by his/her peers, s/he reported more favorable peer interactions, more adaptive perceptions about PE, greater involvement in physical activity, and was rated by the teacher as being highly engaged in his/her PE class.

**Phase 3: Testing the generalizability aspect of validity**

Our focus in phase 3 was on examining the generalizability aspect of validity by considering how score properties and interpretations generalize across tasks and/or time. We sought to address this aspect by having the participants from phase 2 resampled such that they completed the same measures again at a different point in time and in relation to a different focal activity within PE. This resampling allowed us to consider the stability of PAthR scores and structural validity over time, and to examine whether external validity associations observed in phase 3 were similar in direction and magnitude to those observed in phase 2.

**Method**

**Participants and procedure.** Approximately one month after phase 2, 374 ($M_{age} = 13.37, SD = 1.19$) of the phase 2 participants (i.e., a 94% retention rate) completed a survey
that contained the same measures that had been administered at phase 2. Descriptive information indicated that participants who were not present for phase 3 were spread across year groups (i.e., three students missing from grade 7, six students missing from grade 8, five students missing from grade 9, and 11 students missing from grade 10).

**Measures.** All measures were the same as phase 2. Cronbach’s alphas supported the internal consistency of all measures (i.e., PAthR = .93, relatedness support = .93, RISE = .93, relatedness need satisfaction = .95, social anxiety = .92, competence = .90, amotivation = .85, intrinsic motivation = .91).

**Data analysis.** Using the same method as described in phase 2, random effects models were used to test bivariate associations between (a) PAthR reported at phase 2 and phase 3 (representing stability across one month) and (b) phase 3 PAthR scores and all other variables reported at phase 3 (representing concurrent associations). Variable scales were left unadjusted and intercepts and slopes were allowed to vary across classes. As with phase 2, we used the EM procedure to deal with missing data (representing 1.8% of total phase 3 data). In line with Tabachnick and Fidell (2007), we also removed 16 multivariate outliers (i.e., Mahalanobis distance at $p < .001$).

**Results**

We first examined skewness (range -.998 to .08) and kurtosis (range -.69 to 1.24) values for the PAthR items. Only item 4 exceeded the ideal kurtosis value range of -1 to 1. Values outside this range are typical of larger samples and the present sample was sufficiently large so as to avoid underestimates of variance that may otherwise be observed with positive kurtosis (Tabachnick & Fidell, 2007). PAthR item means (prior to being standardized) ranged from 10.25 to 17.59 ($SD$ range 4.09 to 5.63). Corrected item-total correlations for the PAthR measure ranged from .80 to .88, and the deletion of PAthR items would not have improved internal reliability. The ICCs indicated that 23.32% of variability in
PAthR existed at the class level, with item-level ICCs varying between 20.95% and 30.53%. This is nearly twice the amount of variability at class-level compared to phase 2, suggesting class rankings may have become more closely nested across time. Further descriptive information is presented in Table 1.

PAthR assessed in phase 3 was positively associated, $b = .29$ (95% CI: 0.10 to 0.48) with PAthR assessed in phase 2, suggesting the score is relatively stable across time. Bivariate associations between PAthR scores at phase 3 and other constructs measured concurrently are presented in Table 3. The pattern of associations was largely similar to that observed in phase 2. That is, we observed positive associations between PAthR and peer-focused relatedness support, peer-focused RISE, peer-focused relatedness need satisfaction, competence, intrinsic motivation, engagement, and LTPA. Negative associations were observed for PAthR with respect to amotivation and social anxiety. As with phase 2, therefore, analyses revealed that high-reputation students reported more favorable peer interactions, more adaptive perceptions about PE, greater involvement in physical activity, and were rated by the teacher as displaying high behavioral engagement in their class.

**Discussion**

Research in educational settings has shown that individuals’ self-concept, engagement, skill development, and performance may all be shaped by their peer academic reputation (Chen et al., 2010; Gest et al., 2005, 2008; Hughes et al., 2009). To date, though, researchers in the sport and exercise domain have yet to consider if, and how, peer athletic reputation may accompany similar outcomes within an activity context. We proposed that it may be valuable for researchers to have a tool that can be used to assess PAthR, and accordingly, the focus of this manuscript was on (a) the development of an instrument to operationalize PAthR, and (b) the exploration of construct validity evidence for this instrument. With respect to the substantive aspect of validity, item development in phase 1
was guided by the goal of maintaining consistency with existing approaches that have focused on academic competence reputations (i.e., number and phrasing of items, balance of positively and negatively worded items; Gest et al., 2008). Content validity was also attended to as part of the instrument development process by modifying item content to be suitable to the athletic domain and PE context.

In phase 2, we found support for the structural aspect of validity. CFA results generally indicated support for a single-factor (i.e., PAthR) model and scores derived from the PAthR measure displayed an acceptable level of internal consistency. The single-factor structure is consistent with Gest et al.’s (2008) view that although positive and negative aspects of an individual’s reputation may be considered using a bifurcated approach, it is most appropriate to create a composite score that reflects a single reputational index. It is interesting to note, however, that both the item means were higher for the negatively worded (i.e., reverse-scored) items than the positively worded items. This pattern of findings may indicate that participants were either less able or less willing to nominate peers who they believed to have a lower athletic ability, and is consistent with conclusions reported by Gest et al. (2005, 2008). We also acknowledge that a significant chi-square value was observed for the CFA model. Whilst the chi-square value may be considered an overly stringent fit criteria in some cases (Byrne, 2012), the observed value indicates the model may be a close but inexact fit to the data. We recommend that researchers continue to test the factor structure of the proposed measure, such as by using a multilevel CFA to examine support for the single-factor solution. However, we advise investigators that the second-level sample size for a multilevel CFA should ideally exceed that collected in the present work (Maas & Hox, 2005).

Associations found in phase 2 also provided support for the external aspect of validity. Results indicated that individuals whose athletic ability was rated more positively by their peers reported more favorable perceptions about their peer interactions, greater
involvement in PE, and higher levels of participation in physical activity outside of school.

Interestingly, those students who were endorsed as having a strong athletic reputation by their peers were also rated by their teachers as displaying high levels of behavioral engagement in their PE lessons. On the reverse, students who scored relatively low on this peer reputational index appeared to experience relatively high levels of anxiety regarding their peer interactions, and higher levels of amotivation for their PE class. These results are in line with Gest and colleagues’ (2005, 2008) view that reputational indices may be useful markers of individuals’ domain-specific functioning. In particular, the finding that PAthR was positively associated with students’ behavioral engagement (as rated by an external source) indicated support for the proposition that reputation ratings may coincide with functional outcomes in the focal context (Gest et al., 2005, 2008). Parenthetically, the associations between PAthR and related constructs also indicated support for the empirical distinguishability of reputation scores from related peer/competence-based variables.

In phase 3, we addressed the generalizability aspect of validity by re-sampling phase 2 participants at a later time and in relation to a different focal activity. Analyses indicated that, when measured approximately one month apart, PAthR was relatively stable across time points. Peer reputations are theorized to be dynamic, given that peers constantly accumulate information that allows them to make (and revise) social comparisons and judgments regarding their peer-group (Gest et al., 2005). With respect to PE classes, changes in one’s PAthR may be observed given that there are a diversity of focal activities and skills inherent within PE. Given that PAthR items were focused on the sport/activity that was currently being performed in the PE class, it may be that an individual who is viewed more positively in the context of one activity (e.g., soccer) is not necessarily viewed as positively in the context of another activity (e.g., gymnastics, swimming). With that in mind, it would also be useful in future to consider the observable behavior that may help explain changes in peer
reputation over time, and the extent to which modality changes may moderate the generality
of PAthR scores. The process of individuals updating their perception of others’
characteristics (e.g. athletic ability) as they observe their behavior over time may also help
explain why we observed an increase in the percentage of variability in PAthR that was
nested at the class level from phase 2 to phase 3. Although speculative, it may be that this
process is accompanied by an increased nesting effect within classes as perceptions are
refined over time. In phase 3, we also re-examined associations between PAthR scores and
correlate variables. The pattern of associations was largely consistent with that which was
observed at phase 2, and provided support for the external aspect of construct validity.
Notwithstanding some minor differences, we examined a series of important activity- and
peer-related processes as potential correlates, and it appears that there is some temporal
stability in the nomological net with which PAthR aligns.

Although our findings provide some support for the study and measurement of PAthR
in PE settings, we acknowledge that measure development is an iterative process, meaning
that construct validation efforts should be ongoing (Messick, 1995). These efforts may
address some of the limitations of the present work. Participants in the present study were
drawn exclusively from one all girls’ high school and the interpretation of results is limited to
this female sample. Our focus on females was valuable in light of the heightened evaluative
concerns, stronger peer influence, and lower PE and physical activity engagement that has
been reported for females in comparison to males during this developmental period (e.g.,
McKenzie, Marshall, Sallis, & Conway, 2000; Ornelas et al., 2007; Rose & Rudolph, 2006;
Rudolph & Conley, 2005). Nonetheless, we encourage future work that tests the notion that
females may be stereotyped as having lower ability than males in physical activity settings
(Chalabaev, Sarrazin, Fontayne, Boiché, & Clément-Guillotin, 2013). This work could be of
potential use in exploring the antecedents of PAthR, such as examining the extent to which
this reputational marker is predicted by less objective factors (e.g., sex stereotypes) or more
objective factors (e.g., observed behaviors during PE class). This line of research may have
important implications for understanding the extent to which an individual can shape their
own reputation, or the extent to which their reputation is a function of broader societal
expectations.

Another limitation was that the focus of the present work was solely on the PE
context. It would be interesting to expand construct validation efforts to other prevalent
physical activity contexts (e.g., organized sport) so as to further examine the generalizability
aspect of validity. Examining the consequential aspect of validity was beyond the scope of
the present paper but should also be considered going forward. One way to address this
aspect of validity would be to consider the extent to which PAthR precedes, rather than just
correlates with, important physical activity outcomes. In this respect, we endorse research
that seeks to examine not only if PAthR predicts participation, achievement, and effort in
physical activity contexts, but also explores the within-group communication/interaction
processes that accompany PAthR scores. In examining these possibilities, we encourage
investigators to use established and validated measures. In the present work, minor
modifications to existing relatedness support, RISE, and social anxiety measures were made
in the interests of contextual appropriateness. However, these changes also limit claims
relating to the validity interpretation of scores derived from these measures.

Researchers may also wish to draw on theory (e.g., Bandura, 1997; Festinger, 1954;
Lent & Lopez, 2002) so as to guide their investigation of the perceptual pathways, such as
peer RISE, self-efficacy, and reflected appraisals through which PAthR may influence
indices of well-being and achievement. Moreover, in considering the predictive utility of
PAthR, it would be worthwhile to explore the extent to which individuals’ athletic reputations
are concordant (or not) with other aspects of their reputation (e.g., their social standing), and
whether a particular reputational index takes precedence in shaping individual outcomes.

More broadly, it is possible that PAthR has implications for third-party others’ views of the individual. For example, peer-based reputation may shape teacher expectations of students during PE class. It would be interesting to explore any potential bias that stems from non-objective peer-based reputations given the importance of teachers in shaping PE and physical activity outcomes.

From a measurement perspective, it is important to note that additional procedures that could have been implemented during item development were not adopted. For example, rather than generating a pool of new items and having these items reviewed by subject matter experts, the research team modified existing PAR items (Gest et al., 2005, 2008) to suit the PE context. This approach limited our ability to ascertain the extent to which the items are the best possible representations of the PAthR construct. Further to issues of methodology, we encourage researchers to be cognizant of the potential for peer victimization and evaluative concerns when administering PAthR measures. Clearly, it is important to be cautious and provide reassurance when requesting individuals to report on those who do not perform well in PE. In the present investigation, we were careful to inform all students that the measurement of PAthR was in no way intended to inform classmates about the ‘pecking order’ in their PE class. We noted to respondents that although they may have felt that others were not highly competent in the specific activity in question, this did not mean that these students were not competent in any PE activities. In addition, we highlighted that because all students held different qualities and strengths, their perceptions about others’ reputations in PE were not reflective of their reputation in other subjects at school, or their personality or ability in general.

There is a need to further explore the properties of the PAthR instrument; however, the evidence presented in this study provides preliminary support for the assessment of this
reputational marker within PE. The findings indicate that PAthR can be considered a unitary construct that may explain variance in a number of important physical activity outcomes (e.g., PE engagement) and theoretical determinants of these outcomes (e.g., motivation, competence). This was the first investigation to operationalize PAthR within PE, and we encourage future construct validation efforts that advance our understanding of the nature, development, and implications of peer athletic reputation.
References


Table 1. *Descriptive information and factor loadings for the PAthR measure.*

<table>
<thead>
<tr>
<th>Item</th>
<th>Phase 2</th>
<th>Phase 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Factor loading</td>
<td>Mean</td>
</tr>
<tr>
<td>Is good at the sport you are currently doing</td>
<td>.90</td>
<td>9.51</td>
</tr>
<tr>
<td>(Good at reading)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is not good at the sport you are currently doing R</td>
<td>.997</td>
<td>17.34</td>
</tr>
<tr>
<td>(Not very good at reading)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Does activities and skills well in the sport you are currently doing</td>
<td>.83</td>
<td>11.34</td>
</tr>
<tr>
<td>(Almost always knows the right answer)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Does not do activities and drills well in the sport you are currently doing R</td>
<td>.81</td>
<td>17.67</td>
</tr>
<tr>
<td>(Almost never knows the right answer)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean factor loading</td>
<td>.88</td>
<td></td>
</tr>
<tr>
<td>Scale descriptives</td>
<td></td>
<td>13.96</td>
</tr>
</tbody>
</table>

*Note. α = Cronbach’s alpha. R Descriptives presented after reverse-scoring the items. Items indented and in parentheses are from Gest et al.’s (2005) PAR measure and are presented for comparison with PAthR items.*
Table 2. *Phase 2 Results of Hierarchical Linear Models Testing Bivariate Associations with PAthR accounting for Nesting at Class Level*

<table>
<thead>
<tr>
<th>Variable</th>
<th>$b$</th>
<th>95% CI</th>
<th>RE variance (SD)</th>
<th>-2LL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peer-focused relatedness support</td>
<td>0.30*</td>
<td>0.13 to 0.44</td>
<td>0.04 (0.21)</td>
<td>-584.44</td>
</tr>
<tr>
<td>Peer-focused RISE</td>
<td>0.23*</td>
<td>0.17 to 0.30</td>
<td>0.00 (0.07)</td>
<td>-310.28</td>
</tr>
<tr>
<td>Peer-focused relatedness need satisfaction</td>
<td>0.25*</td>
<td>0.10 to 0.41</td>
<td>0.05 (0.22)</td>
<td>-608.73</td>
</tr>
<tr>
<td>Peer-focused social anxiety</td>
<td>-0.26*</td>
<td>-0.35 to -0.15</td>
<td>0.00 (0.00)</td>
<td>-559.62</td>
</tr>
<tr>
<td>Competence</td>
<td>0.62*</td>
<td>0.48 to 0.75</td>
<td>0.02 (0.14)</td>
<td>-540.37</td>
</tr>
<tr>
<td>Amotivation</td>
<td>-0.37*</td>
<td>-0.49 to -0.22</td>
<td>0.04 (0.19)</td>
<td>-551.81</td>
</tr>
<tr>
<td>Intrinsic motivation</td>
<td>0.47*</td>
<td>0.34 to 0.60</td>
<td>0.03 (0.16)</td>
<td>-573.52</td>
</tr>
<tr>
<td>PE engagement (teacher rating)</td>
<td>0.59*</td>
<td>0.39 to 0.81</td>
<td>0.16 (0.40)</td>
<td>-503.04</td>
</tr>
<tr>
<td>LTPA</td>
<td>9.44*</td>
<td>5.67 to 12.66</td>
<td>6.55 (2.56)</td>
<td>-1846.52</td>
</tr>
</tbody>
</table>

*Note.* PAthR = peer athletic reputation, using standardized Z scores. LTPA = self-reported leisure-time physical activity. $b$ values marked with an asterisks have a 95% CI not encompassing zero. RE = random effect. -2LL = -2 log likelihood.
### Table 3. Phase 3 Results of Hierarchical Linear Models Testing Bivariate Associations with PAthR accounting for Nesting at Class Level

<table>
<thead>
<tr>
<th>Variable</th>
<th>$b$</th>
<th>95% CI</th>
<th>RE Variance (SD)</th>
<th>-2LL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peer-focused relatedness support</td>
<td>0.39*</td>
<td>0.22 to 0.55</td>
<td>0.05 (0.23)</td>
<td>-538.55</td>
</tr>
<tr>
<td>Peer-focused RISE</td>
<td>0.28*</td>
<td>0.18 to 0.40</td>
<td>0.03 (0.17)</td>
<td>-319.13</td>
</tr>
<tr>
<td>Peer-focused relatedness need satisfaction</td>
<td>0.50*</td>
<td>0.28 to 0.72</td>
<td>0.11 (0.32)</td>
<td>-517.26</td>
</tr>
<tr>
<td>Peer-focused social anxiety</td>
<td>-0.35*</td>
<td>-0.50 to -0.21</td>
<td>0.04 (0.19)</td>
<td>-489.28</td>
</tr>
<tr>
<td>Competence</td>
<td>0.59*</td>
<td>0.45 to 0.74</td>
<td>0.01 (0.11)</td>
<td>-514.77</td>
</tr>
<tr>
<td>Amotivation</td>
<td>-0.49*</td>
<td>-0.63 to -0.36</td>
<td>0.01 (0.11)</td>
<td>-513.93</td>
</tr>
<tr>
<td>Intrinsic motivation</td>
<td>0.68*</td>
<td>0.45 to 0.90</td>
<td>0.13 (0.36)</td>
<td>-533.26</td>
</tr>
<tr>
<td>PE engagement (teacher rating)</td>
<td>0.63*</td>
<td>0.43 to 0.83</td>
<td>0.13 (0.37)</td>
<td>-464.96</td>
</tr>
<tr>
<td>LTPA</td>
<td>8.68*</td>
<td>1.42 to 16.22</td>
<td>53.11 (7.29)</td>
<td>-1814.72</td>
</tr>
</tbody>
</table>

*Note. PPathR = peer athletic reputation, using standardized Z scores. LTPA = self-reported leisure-time physical activity. $b$ values marked with an asterisks have a 95% CI not encompassing zero. RE = random effect. -2LL = -2 log likelihood.*