A predictive validity study of the Learning Style Questionnaire (LSQ) using multiple, specific learning criteria

F.R. Kappe, L. Boekholt, C. den Rooyen, H. Van der Flier

A R T I C L E   I N F O

Article history:
Received 2 April 2008
Received in revised form 25 March 2009
Accepted 6 April 2009

Keywords:
Learning styles
Academic achievement
Learning outcomes
GPA
Predictive validity

A B S T R A C T

Multiple and specific learning criteria were used to examine the predictive validity of the Learning Style Questionnaire (LSQ). Ninety-nine students in a college of higher learning in The Netherlands participated in a naturally occurring field study. The students were categorized into one of four LSQ dimensions, namely, Activists, Theorists, Pragmatists, and Reflectors, and they were also graded on student GPA as a criterion measure. The purpose of the present study was to provide a more accurate test of the learning style model by pairing the various learning styles with multiple and specific performance criteria suited to each learning style preference.

One model of learning styles that has generated a significant amount of research is that of David Kolb (1984). While teaching management students he noticed that some students preferred learning through experiences whereas others preferred the traditional classroom lecture. His subsequent theory of experiential learning proposed that, while learning, people resolved conflicts between a) active experimentation and b) reflective observation along one axis and between c) concrete experience and d) abstract conceptualization along another axis. His model yielded four quadrants and he stated that, over time, people developed learning style preferences that can be categorized into one of the four quadrants. Kolb developed the Learning Style Inventory (LSI) to measure peoples’ individual learning styles. By knowing the learning styles of their students and by creating learning environments matched to those learning styles, educators could enhance learning. Whereas commercially popular pedagogical tools have been generated by Kolb’s work, empirical support for construct and predictive validity have been lacking (Coffield et al., 2004).

Difficulty with some of Kolb’s theoretical ideas and low face validity for the LSI prompted Honey and Mumford (1986) to develop their own learning style theory as well as a new measure, namely, the Learning Style Questionnaire (LSQ). The LSQ also identified four types of learners: Activists, Theorists, Pragmatists, and Reflectors. Activists are individuals who enjoy new experiences, tend to make decisions intuitively, but who dislike structured procedures. Theorists focus on ideas and systemic logic and are distrustful of intuition and emotional involvement. Pragmatists like practical, down to earth approaches and debate, but tend to avoid reflection and deep levels of understanding. Reflectors observe and describe processes, try to predict outcomes and try to understand meaning.

Despite its commercial success, empirical support for the LSQ has been mixed. Allinson and Hayes (1988) concluded that the temporal stability and internal consistency of the instrument were well established and provided some evidence of construct validity but were unable to find concurrent or predictive validity. Whereas some researchers have reported good test–retest reliabilities (Duff, 2000; Honey & Mumford, 2000; Swailes & Senior, 1999; Veres, Sims, & Locklear, 1991), others have questioned the internal consistency of the LSQ or reported low test–retest values (Jackson & Lawty-Jones, 1996; Ruble &
Stout, 1993; Sims, Veres, & Shake, 1989). Duff and Duffy (2002) employed exploratory and confirmatory factor analysis but failed to find four separate learning styles and reported that learning style was a weak predictor of academic performance. Price and Richardson (2003) used the LSQ to predict student performance, study techniques, and recall processes, but failed to find significant results. Finally, Zwanenberg, Wilkinson, and Anderson (2000) were unable to predict learning outcome scores for any of the four LSQ subscales and surmised that the use of generalized, instead of specific, criterion measures might have accounted for their lack of findings.

Most predictive validity studies in educational settings have used generalized criterion measures (O’Conner & Paunonen, 2007), namely, grade point average (GPA). An overall indicator such as GPA typically reflects arithmetically averaged components of performance and one could well have averaged out meaningful variance with such a criterion. It is quite possible that the weak and nonsupportive findings mentioned above were attributable to the use of insensitive criterion measures. It is for this reason that O’Conner and Paunonen (2007) have called for decomposing the broad criterion measures, such as overall GPA, into more specific components. The purpose of our study was to test learning style theory by using a number of specific, different criteria that could be logically related to the four learning styles measured by the LSQ.

Honey and Mumford (2000) have developed a table in which they describe how individuals who score high on each of the four different learning styles prefer to learn. This table served as the basis for linking the learning styles to the multiple performance criteria available to us in this study. For example, the table states that Activists prefer action learning, job rotation, role playing, business game simulations, discussion in small groups, training others, and outdoor activities. We therefore hypothesized that Activists would score high on a measure of on-the-job training as well as score high on a measure of practical skills training. Similarly, Reflectors, who prefer listening to lectures, reading, e-learning, and self-directed learning should score high on a criterion that measured attentiveness to lecture material. Finally, Pragmatists, who prefer action learning, discussion in small groups, problem solving workshops, and group tasks where learning is applied, should score high on a criterion that measured how well students, working in groups, developed in-depth programs that met practical business needs.

1. Method

1.1. Participants and setting

Undergraduate students in a professional school of higher educational learning, majoring in human resource management (HRM), participated in a naturally occurring field study. Students studied such HRM issues as how to restructure the long term unemployed into the workforce, or how to help employees plan their current careers, as well as the legal steps involved in employee termination. The current HRM program uses a variety of learning environments such as classroom tuition, group projects, and skills training. Also a large amount of time is devoted to on-the-job training in an actual business setting during the second, third and fourth year, and students are required to write a thesis in the last six months of their study.

Students had finished their secondary school education and were generally in the 18 to 22 year age bracket. Data were collected from four cohorts of students (2000–2004, n = 34; 2001–2005, n = 44; 2002–2006, n = 57; and 2003–2007, n = 39). There were a total of 174 students, 99 of whom completed the entire HRM program, 21 of whom started but did not finish the program, and the remaining 54 of whom either entered the program in midstream, having transferred from other schools, or had missing data values. The final sample consisted of the 99 matriculated students and of these, 31% of the participants were male and 69% were female.

A power analysis was conducted to determine the probability of finding an expected effect size. With a posited medium effect size of .30 and an alpha level of .05, a sample size of 99 yields a sufficient power value of .85 (Cohen, 1975).

1.2. Measures

1.2.1. Predictor

The Learning Style Questionnaire (LSQ) was chosen as a measure of learning style because it, a) has shown face validity, b) has been popular commercially, and c) because it was originally developed for managers and should therefore generalize to the participants in our study, namely, students who were being trained to become human resource managers. All students completed the LSQ twice; once at the end of their first year (T1) and then again at the end of their third year (T2), both times as part of an Assessment Center. The T1 score was used to compute the LSQ score for each student. LSQ test–retest reliability scores were based on correlations between T1 and T2 scores.

The LSQ (1992) consists of four learning style subscales: Activist, Reflector, Theorist, and Pragmatist; each containing 20 items for a total of 80 items. Respondents either agreed or disagreed with the statements and subscale scores were determined by summing the positive responses, thus yielding a theoretical range of 0 to 20. As noted by Honey and Mumford, the LSQ has been translated into many languages and the Dutch language version was used in this study.

1.2.2. Criterion

Academic achievement data were obtained from the student administration office. Students received grades ranging from one to ten for each assessment they completed with a grade of six representing a pass. However, as is typical in the Dutch education system, students who failed an assessment on their first attempt were permitted to try again. Because first-try failed grades were not recorded by the school, only grades of six and above were available to the researchers.

Five criterion measures were obtained for each student, one each for: classroom lectures, skills training, group projects, on-the-job training, and written thesis. Each criterion score represented an average of multiple assessments throughout the student’s four year curriculum. Except where noted, students were appraised by their teachers. Atteniveness to classroom lectures and assigned textbook readings was measured with multiple choice and essay exams. Skills training measured how well students mastered such skills as negotiating and debating, as well as curriculum specific skills such as how to conduct an employment interview. For group projects, students worked in groups of five, each typically lasting ten weeks, to develop HRM programs (e.g. selection and recruitment) for businesses. For on-the-job training, students worked as junior employees in actual business settings in curriculum relevant areas such as recruiting and applying government employment regulations. Assessment scores were obtained from discussions between the workplace supervisors and the students’ mentors. Finally, written theses, typically 30 pages in length, were required in which students discussed practical business problems, often stemming from their on-the-job training (OJT) experiences, and for which they offered potential solutions. The students’ mentors, as well as independent judges, rated the written theses and oral presentations thereof, and, after discussion, agreed to overall scores.

The number of independent assessments for each student for each criterion was as follows: classroom lectures (25), skills training (10), group projects (10), on-the-job training (3), and thesis (2). In the same way that the mean of a distribution of scores is a more stable index than is a single score from that distribution, obtaining multiple assessment measures provided additional stability for each criterion score.
2. Results

Descriptive statistics for all predictor and criterion variables are listed in Table 1. Tests for skewness, kurtosis, and Kolmogorov–Smirnov normality were conducted to ascertain whether the criterion data were normally distributed. Results showed that indices of skewness and kurtosis were <1 for all criteria except thesis. Furthermore, Kolmogorov–Smirnov tests for normality revealed that three of the five criteria were normally distributed, the exception being on-the-job training ($p = .001$) and thesis ($p = .014$).

Students were classified into one of the four learning style subgroups according to within-student rank-order scores. Thus if a particular student had the following scores: 7/activist, 13/reflector, 17/theorist, and 12/pragmatist, then he or she would be classified as being a theorist. Results indicated that 20% of the students preferred an activist learning style, 46% a reflector learning style, 17% a theorist learning style and another 17% a pragmatist learning style. We can report positive findings concerning test–retest reliabilities for the LSQ. Test–retest reliabilities for the four subscales, over the two year period, were: Activists ($r = .70$), Reflectors ($r = .63$), Theorists ($r = .50$), and Pragmatists ($r = .46$). These values, taken in context, are quite respectable because test–retest values above $r = .50$ are normally considered acceptable. Whereas most researchers report acceptable overall test–retest reliabilities ranging from .60 to .90 across a two to eight week time period, with Duff (2000) extending that to one year and reporting an overall value of .74, we are able to report values ranging from $r = .46$ to $r = .70$ over a two year period.

The main hypothesis of this study was tested with bivariate correlations and results are presented in Table 2. Contrary to expectations, none of the correlations between learning styles and criterion measures were significant. In an effort to increase the power beyond 90%, the analyses were rerun, adding the 21 non-matriculating students who had dropped out of the program, so that total $n = 120$, however, none of the correlations were significant.

Because there were no significant predictive validity results, we tried several tests using mismatched conditions to determine whether the LSQ would differentiate among groups. Several researchers have recommended testing the LSQ using mismatched conditions (e.g. Stellwagen, 2001). Results showed that Activists did not score higher than a combined Reflector/Theorist group on the criterion of on-the-job training, $t = -.262$, $df = 78$, and $p > .05$; nor did they do so on the skills training criterion, $t = .066$, $df = 78$, and $p > .05$. Reflectors did not score higher than a combined Activist/Pragmatist group on the classroom lecture criterion, $t = .928$, $df = 82$, and $p > .05$; nor did Theorists score higher than a combined Activist/Pragmatist group on this criterion, $t = -1.902$, $df = 51$, and $p > .05$. Finally, no significant results were found, on the criterion of classroom lecture, when students who scored high on either Reflectors or Theorists were combined into one group and compared to a group of students who scored low on either of these orientations, $t = -.138$, $df = 97$ and, $p > .05$.

3. Discussion

The most important finding of this study is, in essence, a non finding. Despite the use of specific performance criteria tailored to the four learning styles, we are unable to predict student achievement scores using LSQ sub scores. The lack of findings cannot be attributed to unreliability in the LSQ, nor to a lack of power in this research.

Learning style theory is a subset of personality theory which, as a theory of human behaviour, fails to receive sufficient empirical support. Bandura (1986) roundly criticizes personality theories saying that numerous studies have shown that measures of personality traits correlate weakly with social behaviour in different settings. Rather, people change their behaviour depending on the functional value of that behaviour in a particular setting. Along the lines of Bandura’s Social Cognitive Learning Theory, Price and Richardson, commenting on their inability to predict student achievement with the LSQ, conclude that, “...tests of generalized individual differences are inappropriate for understanding performance in task-specific and context-specific situations” (2003, p. 294).

That we are unable to report positive findings is perhaps not surprising. Smith, Sekar, and Townshend (2002) provide a review of the literature regarding the matching hypothesis and report that for every study that supports the hypothesis there is a study which rejects it. Such a difference in findings can sometimes be explained by a hidden moderator variable and were one to conduct a meta-analysis, in which studies are grouped according to levels of a moderator variable, one might explain why some studies support whereas others reject a particular finding (Wiersma, 1992). Generalized GPA scores versus specific performance scores may be such a moderator and further research might benefit by grouping studies according to how specific the criteria are, however results of our study are not encouraging.

Limitations in the current study may account for the lack of positive findings and considering our claim of non findings it is important to consider any limitations carefully. An argument can be made about restriction of range in the performance criteria but this argument becomes debatable upon closer review. Theoretical scores on our criteria can range from six to ten, but we acknowledge that a ten is almost never awarded and a nine is also quite rare. Thus the practical range is from six to eight. However, this is not all that different from, for example, the American higher educational system. Grades in the U.S. can range from ‘A’ to ‘F’, in which an A is worth four points and an F is worth zero. A straight ‘A’ student would receive a GPA of 4.0. Here too, the actual grade distribution is restricted because students often drop courses that they are in danger of failing, and grade inflation, embodied in the gentleman’s ‘C’ for sub par work, is common on college campuses. One is also left essentially with a three point scale from A to C. Thus, from a practical point of view, it seems that future field tests of the learning style model may very well be confronted with this fact.

From a methodological viewpoint, restriction of range is a concern because it restricts the amount of variance that can be predicted and therefore the size of obtainable correlations. This we acknowledge, however we wish to note that we obtained multiple assessment
measures per criterion which lessens the ‘noise’ around each individual criterion score. Basing a classroom lecture on 25 assessments across a four year curriculum yields a much more stable criterion score than one based on one or two assessments because single assessments contain too much error variance. By the same token, a distribution of criterion mean scores reduces the total observable variance but the variance which remains is less affected by error variance.

Honey and Mumford have responded to the lack of empirical support for their theory by stating that the LSQ was never meant to be a psychometric instrument but rather a simple self-developmental tool that makes people aware of how they learn. That is like designing an aircraft in a wind-tunnel according to certain principles of aerodynamics only to discover that the aircraft cannot fly. Surely the theoretical formulations (in both instances) are highly suspect. If the only purpose of the LSQ is as a stimulus to focus peoples’ attention on learning, then any simple theory that looks interesting to people will do. However if the purpose is to increase learning efficiency, then we should continue to look for a theory that will truly provide people with the wings to fly. As Kurt Lewin noted, there is nothing as practical as good theory.

Acknowledgements

The authors acknowledge the reviewers for their helpful suggestions on earlier versions of this article and also Uco Wiersma for helpful suggestions at the finishing phase. The present investigation is part of a larger PhD study funded by INHOLLAND as well as the LTP institute.

References


