Examining the Effect of Different Measurements of Learning Success in Technology-mediated Learning Research

Research-in-Progress

Andreas Janson¹
Matthias Söllner¹,²
Philipp Bitzer¹
Jan Marco Leimeister¹,²

¹ Kassel University
Information Systems
Pfannkuchstr. 1, 34121 Kassel, Germany
[andreas.janson, soellner, bitzer, leimeister]@uni-kassel.de

² University of St. Gallen
Institute of Information Management
Mueller-Friedberg-Str. 8, 9000 St. Gallen, Switzerland
[matthias.soellner, janmarco.leimeister]@unisg.ch

Abstract

This research-in-progress paper examines the effects of different measurement methods for learning success with respect to the empirical evaluation of technology-mediated learning (TML). We argue that the use of self-reported data for the major dependent variable of TML, learning success, is insufficient and a major validity threat of past research results; thus, we examine the effect of employed measurement methods in TML against the background of common method variance. We are currently conducting a study on the antecedents of learning success measured by three different approaches that include self-reported learning and objective learning success. By analyzing the data, we are able to investigate how different measurement approaches to learning success impact research findings. Our contribution to theory and practice is an assessment of the validity of self-reported learning success measures and the impact of different measurement approaches for the relationships in a TML model.

Keywords: Common Method Variance; Technology-mediated Learning; Learning Success; Self-reported Learning
**Introduction**

Learning and educational scenarios are widely influenced by the use of technology. These scenarios are typically referred to as technology-mediated learning (TML). E-learning often refers to the same phenomenon and is widely recognized as an important area in IS research. Apart from classical e-learning, the aim of TML is the integration of both face-to-face as well as IT-based learning activities. We therefore define TML as “environments in which the learner’s interactions with learning materials (readings, assignments, exercises, etc.), peers, and/or instructors are mediated through advanced information technologies” (Alavi and Leidner 2001, p. 2). For the empirical evaluation of TML success, many researchers focus on the learning outcomes of TML as dependent variable (e.g., Bitzer et al. 2013; Gupta and Bostrom 2013, 2009). However, learning outcomes, such as learning success, can be measured in very different ways, and their appropriateness has caused an extant discussion in the literature (Benbunan-Fich 2010; Sitzmann et al. 2010; Ross 2006).

Many papers rely on self-reported learning success of TML participants, measured by multiple items using a Likert response format (Alavi 1994; Benbunan-Fich 2010). Critics of this approach have raised concerns, especially regarding the participants’ ability to correctly judge their own, perceived learning success (Benbunan-Fich 2010; Ross 2006; Falchikov and Boud 1989). These concerns are further supported by recent methodological studies pointing out that common method variance (CMV) is a major validity threat to relationships between two variables: if the data for both variables stem from the same recipient and especially if the variables are captured using perceptually anchored scales (see section 2 for further information on the different types of scales) (Podsakoff et al. 2003; Sharma et al. 2009). Since many studies that have contributed to the theoretical foundations on antecedents of learning success have followed the exact approach criticized by Sharma et al. (2009), the reliability of the theoretical foundations is in question (Benbunan-Fich 2010).

We aim to contribute to the assessment of the reliability of the current theoretical foundations by answering the following two research questions:

1. How valid is the self-reported learning success of TML participants compared to their objective learning success?

2. What is the impact of different learning success measurement approaches for measuring the relationships between learning success and its antecedents?

To answer the two outlined research questions, we are currently conducting a study on antecedents of the learning success of TML participants, including three different approaches for measuring learning success (see section 2 for further information). Thereby, we contribute to IS theory and practice by assessing the validity of self-reported learning success instruments and by investigating the impact of different types of measurements on the relationships between learning success and its antecedents. Based on these results, we are able to provide recommendations for both IS researchers and practitioners on how they should measure learning success, and we evaluate the reliability of the current theoretical foundation.

The remainder of this research-in-progress paper is organized as follows. We first present the motivation of our research, the theoretical background on measuring learning success and CMV, and develop the hypotheses for our study. We then present details regarding the theoretical model and the methodology of our study which is currently running. The research-in-progress paper closes with limitations and future research directions as well as a presentation of the current progress, the next steps and expected contributions once our study is completed.

**Theoretical Background and Hypotheses Development**

As we have indicated, our research-in-progress paper focuses on the evaluation of the effect of different measurement approaches for learning success in the context of TML. First, we briefly introduce learning outcomes – in our case learning success – as our phenomenon of interest as well as the recent discussion about the suitability of different measurement approaches in the literature. In a next step, we present the impact of CMV on empirical research findings and show why CMV matters in terms of TML research.
In a broad sense, learning outcomes “represent the goal assessment or measures for determining the accomplishment of learning goals” (Gupta and Bostrom 2009, p. 713). There are several categories for learning outcomes; e.g., Gagné (1985) has identified: intellectual skills, motor skills, verbal information, cognitive strategy, and attitude as learning outcomes. The actual learning process in gaining certain learning outcomes includes, as a prerequisite, the acquisition of knowledge and the capability to perform and take effective actions on the basis of this knowledge (Alavi and Leidner 2001). However, Alavi and Leidner (2001) state that knowledge and the capability to take an action are not directly observable and are therefore latent constructs. In contrast, the action and performance itself can be measured, as it manifests, for example, in test results. In addition, the measurable learning outcomes differ according to the several types of outcomes described above. Therefore, Alavi and Leidner (2001) have called for caution when investigating learning outcomes and approaches to measure learning outcomes such as learning success.

TML research often measures learning outcomes such as learning success on the basis of self-assessment (Sitzmann et al. 2010). We view this concept as synonymous to self-reported learning success. However, it is questionable whether this particular construct is measurable with self-assessment techniques due to validity threats (Falchikov and Boud 1989). Ross (2006) purports that self-reported learning success is usually higher than objective learning outcomes, e.g., as rated by teachers. However, it should be noted that exceptions have been reported in which the self-reported learning success was lower than the objective (e.g., Aitchison 1995). We thus do not want to predict any direction of effect. On the basis of these arguments, we hypothesize:

**H1: A learner’s self-reported learning success differs from their objective learning success.**

Connected to our first and overall hypothesis is the statistical concept of CMV, which describes the bias of measurement caused by similar sources for endogenous and exogenous variables (Podsakoff et al. 2003; Sharma et al. 2009). These biases are measurement errors that arise from construct-level relationships between the calculated variables which result in inflated correlations (Sharma et al. 2009). Therefore, CMV threatens the validity of findings concerning the relation of variables, and consequently poses a great problem. The measurement errors consist of a random and a systematic component, both of which are considered to be problematic. While the random measurement error is comparably harmless, the systematic measurement error provides an alternative explanation for the observed relationships, thus posing a serious problem. Systematic measurement errors result from method variance that arises from a variety of sources, e.g., the content of a specific item, scale type, response format, and the general context (Podsakoff et al. 2003). At the same time, these mentioned factors have an increasing effect on CMV. Especially measuring independent and dependent variables at the same time poses a larger CMV-based potential validity threat. Further, method effects can be interpreted in terms of response biases, such as halo-effect, social desirability, acquiescence, leniency effects, or yea- and nay-saying (Podsakoff et al. 2003). Hence, systematic error variances influence empirical results through misleading conclusions.

Several sources and research settings foster CMV (Sharma et al. 2009; Podsakoff et al. 2003). The sources can be divided in method effects caused by a common source or rater and other method effects resulting from the measurement items themselves or the context of measurement. In the first case, method effects result from self-reporting because respondents and predictors are the same persons: they try to maintain consistency between their cognitions and attitudes, and hope to achieve rational answers. Another source is socially accepted behavior, which means that people try to adapt to culturally acceptable and appropriate behavior.

Apart from the aforementioned sources of CMV, there are several other potential sources/method characteristics (Podsakoff et al. 2003). Four groups of CMV can be differentiated: source, item characteristics, item construct, and measurement context (Sharma et al. 2009; Podsakoff et al. 2003). Item characteristics result from common scale formats to measure dependent as well as independent variables. This has an increasing effect on CMV. When both variables are measured by employing items that are abstract, ambiguous, and complex, the item characteristic effect might arise. CMV increases when both variables are measured using abstract items, if respondents are required to engage in more cognitive processing. These are item construct effects. The measurement context can be influenced by measures of dependent and independent variables that occur simultaneously, consequently introducing a covariance in the observed relationships. A temporal separation between the measures would reduce the effect of CMV.
With respect to the described four groups of CMV, four method characteristics can be derived from this: data sources, scale format, item abstractness, which belong to the instruments, as well as temporal separation between measurements describing the manner of instruments administration. More specifically, we investigate in our context CMV that arises through the use of the instrument and measurement methods. In research, the four measurement methods used are: system-captured, behavioral continuous, behaviorally anchored, and perceptually anchored. A detailed description of the measurement methods is given in the following table with reference to learning success measures (Sharma et al. 2009):

<table>
<thead>
<tr>
<th>Method Type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>System-captured</td>
<td>Actual data are obtained from historical records and other objective sources, including usage records captured by a computer system. Example item: Computer-recorded percentage score.</td>
</tr>
<tr>
<td>Behavioral continuous</td>
<td>Items refer to specific behaviors or actions people have carried out. Responses are generally captured on a continuous open-ended scale. Example item: Self-reported percentage score.</td>
</tr>
<tr>
<td>Behaviorally anchored</td>
<td>Items refer to specific actions people have carried out. Responses are generally captured on scales with behavioral anchors, such as “Never - More than once a day.” Example item: /</td>
</tr>
<tr>
<td>Perceptually anchored</td>
<td>Items that capture responses generally on Agree-Disagree Likert scales or on semantic differential scales. Example item: “My technical knowledge has definitely increased.”</td>
</tr>
</tbody>
</table>

Considering the above described measurement methods, IS research is affected by the use of perceptually anchored measurement of items and latent constructs. For example, the Technology Acceptance Model (Davis 1989) is a typical model that is based on perceptually anchored measures, both independent and dependent variables. However, CMV strongly differs considering the use and combination of different measurement methods. Table 2 ranks the expected CMV for a specific method-method pair for our special case of TML, where the independent variable is usually measured on a perceptually anchored scale and the dependent variable is measured with one of three methods (see Table 1 and Sharma et al. 2009). However, we have excluded behaviorally anchored measures since we were not able to identify suitable behaviorally anchored items to capture learning success. Therefore we are not able to provide any example item in the table above.

When the dependent variable is measured using perceptually anchored scales, this particular method-method pair is most susceptible to CMV. The ranking is associated with the conclusion from Podsakoff et al. (2003, p. 885): “Method biases are likely to be particularly powerful in studies in which the data for the predictor and criterion variables are obtained from the same person, on the same measurement context, using the same item context and similar measurement context, using the same item context and similar item characteristics” (Podsakoff et al. 2003). Table 2 therefore depicts the rank order for the expected CMV considering the investigated method-method pairs we want to examine in our study.
Considering the depicted method-method pairs, correlations of perceptually anchored measures with objective, system-captured measures are least susceptible to CMV, because the objective measures come from a different source, with a different response format, are factual and verifiable, and they do not depend on the cognitive processing of the respondents with regards to their learning success (Sharma et al. 2009). In contrast, behavioral continuous measures are more susceptible to CMV since they are also self-reported. However, the different and less abstract response formats in comparison to perceptually anchored measures should reduce CMV (Sharma et al. 2009), for instance exam results on a percentage scale may be less abstract than perceptually anchored scales for a student. Our last method-method pair under investigation considers the correlation of perceptually anchored scales for both independent and dependent variables which are both self-reported, employ similar scale anchors, and are more abstract, inducing in consequence a high amount of CMV. On the basis of these considerations, we hypothesize:

H2a: The strength of the relationships between learning success and its antecedents differs, comparing system-captured and behavioral continuous measurement approaches for learning success.

H2b: The strength of the relationships between learning success and its antecedents differs, comparing system-captured and perceptually anchored measurement approaches for learning success.

H2c: The differences observed between the system-captured and perceptually anchored measurement approaches are higher than the differences observed between the system-captured and behavioral continuous measurement approaches.

Concerning our hypotheses, we highlight that we do not want to predict a direction of the effects induced by CMV, since Cote and Buckley (1988) note that method effects can both inflate as well as deflate the observed relationships.

**Theoretical Model of Technology-mediated Learning**

To test the hypotheses of our study, we base our research model on a recent TML quality model that considers structural input factors for TML design, learning process variables, and learning outcomes in line with recent TML research (Gupta and Bostrom 2009). We refer to the already published results of the
model evaluation for a detailed theoretical development of the model (Bitzer et al. 2013). The model is shown below.

![Research Model (on the basis of Bitzer et al. 2013)](image)

We have chosen this model because it considers different antecedents of learning success, namely predisposition of the service recipient (i.e., the learner), the TML process quality, as well as the structural quality of the employed learning materials. It is therefore suitable to investigate how a different measurement approach for learning success relates to very different antecedents and their relationships in a structural model. Therefore, we are able to draw comprehensive conclusions on how the structural relationships differ with respect to different learning success antecedents and thus derive implications whether research is over- or underestimating the impact of learning success antecedents when a certain measurement approach is employed. Moreover, as depicted in Figure 1, by means of our model, we intend to examine how different measurement approaches influence the values of learning success and which measurement approach is most suitable to capture the actual, objective learning success.

**Methodology**

With regards to our hypothesis testing and belonging data collection, the study is already running. It has been conducted with participants of software trainings. In collaboration with one of our partners, we
could access the participants of a series of nine-day software trainings focusing on established ERP software. The participants completed two questionnaires including the items related to our five constructs (see Table 3).

<table>
<thead>
<tr>
<th>Latent Construct</th>
<th>Latent Construct Type</th>
<th>Sub-constructs</th>
<th>Sub-construct Type</th>
<th>Items</th>
<th>Literature Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-disposition of Service Recipient</td>
<td>Formative</td>
<td>Previous Knowledge</td>
<td>Reflective</td>
<td>2</td>
<td>(Bitzer et al. 2013)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Motivation</td>
<td></td>
<td>3</td>
<td>(Colquitt et al. 2000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Technology Readiness</td>
<td></td>
<td>4</td>
<td>(Parasuraman 2000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Self-Regulated Learning</td>
<td></td>
<td>4</td>
<td>(Colquitt et al. 2000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expectations</td>
<td></td>
<td>3</td>
<td>(Bitzer et al. 2013)</td>
</tr>
<tr>
<td>Structural Quality</td>
<td>Formative</td>
<td>Trainer Characteristics</td>
<td>Reflective</td>
<td>9</td>
<td>(Arbaugh 2001; Kim et al. 2011; Parasuraman et al. 1988)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Offline Learning Environment</td>
<td></td>
<td>3</td>
<td>(Ma et al. 2005; Parasuraman et al. 1988)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IT-System Quality</td>
<td></td>
<td>5</td>
<td>(Delone and McLean 2003)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Learning Materials</td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>TML Process Quality</td>
<td>Formative</td>
<td>Interactivity</td>
<td>Reflective</td>
<td>3</td>
<td>(Siau et al. 2006)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IT-Process Support</td>
<td></td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Learning Group</td>
<td></td>
<td>3</td>
<td>(Bitzer et al. 2013)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quality of Exercises</td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Training Process</td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Transparency</td>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Learning Success (Perceptually Anchored Scales)</td>
<td>Reflective</td>
<td>-</td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Learning Success (Behavioral Continuous Scales)</td>
<td>Reflective</td>
<td>-</td>
<td></td>
<td>-</td>
<td>1 Own Indicator</td>
</tr>
<tr>
<td>Learning Success (Objective Measurement)</td>
<td>Reflective</td>
<td>-</td>
<td></td>
<td>-</td>
<td>Training Provider</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>Reflective</td>
<td>-</td>
<td></td>
<td>-</td>
<td>2 (Arbaugh 2001)</td>
</tr>
</tbody>
</table>
The first questionnaire captures the predisposition of the service recipient, whereas the other four constructs are captured in the second questionnaire. The trainers were asked to hand out the first questionnaire on the first or second day of each course. The second questionnaire was completed during one of the last two days of each course. To date, 161 participants have completed both questionnaires that could be used for our evaluation. On average, the participants were 25.57 years of age and 54 were females. Finally, our partner of the software training will provide us with the objective learning success data (%-score of each participant). The two self-reported measurement approaches for learning success were combined in the questionnaire (we did not implement a behaviorally anchored approach since we could not find suitable items). We were thus able to evaluate our research model and the proposed hypotheses.

To evaluate the research model, we use the variance-based partial least squares (PLS) approach (Chin 1998; Wold 1982). As a tool of analysis, we will use SmartPLS 2.0 M3 (Ringle et al. 2005). We decided to rely on hierarchical latent variables (type II) (Jarvis et al. 2003; Ringle et al. 2012) for the pre-disposition of the service recipient, structural quality, and TML process quality, as well as on reflective measurement models for the two learning outcomes, satisfaction, and learning success. The latter is measured with the three different measurement approaches discussed in our paper: perceptually anchored and behavioral continuous scales, both measuring self-reported learning success, as well as an objective measurement of learning success. To evaluate the hierarchical latent variables, we will use the two-stage approach, because our structural model contains an endogenous, hierarchical latent variable (TML process quality) (Ringle et al. 2012; Becker et al. 2012). In the first stage, latent variable scores are obtained for the lower order components (e.g., trainer characteristics as a lower order component of the structural quality) by using the repeated indicator approach (Wetzels et al. 2009). In the second stage, the latent variable scores are used as manifest variables for the higher order components (Becker et al. 2012; Ringle et al. 2012). In contrast, a pure repeated indicator approach would be not suitable, since all path relationships to the endogenous, second-order construct of TML process quality would be approximately zero and non-significant (Ringle et al. 2012).

Finally, to evaluate the impact of different measurement approaches for learning success, we evaluate the proposed hypotheses and different method-method pairs with a multi-group analysis (Hair et al. 2014). Thus, we evaluate if there are significant differences in path estimates between the measures for the dependent variables (Gefen et al. 2011).

Limitations and Future Research

Our research-in-progress paper is not without limitations. In addition to the discussion of the measurement of learning success, the concept of learning success itself is still vague and differs in scope and level (Benbunan-Fich 2010). Referring to our objective data, we will be provided with objective data that captures learning success shortly after the training. However, research also suggests the consideration of the long-term impact of training measures (Gupta and Bostrom 2009). Therefore, an important addition to learning outcomes would be constructs such as delayed task performance in a work context (Santhanam and Sein 1994; Sein and Bostrom 1989; Sein et al. 1999), which should be considered in upcoming studies and future research. This relates also to a possible reconceptualization of learning success itself. We therefore recognize that both perceived as well as objective measures for learning success are both important variables (Benbunan-Fich 2010). Future research should therefore better define what perceived learning success actually is and why it may differ from objective learning success beyond any methodological issues discussed in this paper.

Current Progress, Next Steps and Expected Contributions

We have preliminarily finished our theory development, hypotheses, and research design, and are currently in the final stage of our data collection. The trainings are now finished and all self-reported data have already been collected. However, we are waiting for the objective learning success data provided by the software training institution. Therefore, our study is still a research-in-progress. Next steps after receiving the data will include the evaluation of the proposed hypotheses and the interpretation of our results.
Our paper contributes to the body of IS theory with two major outcomes: first, with the assessment of the validity of self-reported learning success instruments and by investigating the impact of different types of method-method pairs. We thus gain insights on how different measurement methods and method-method pairs impact the observed relationships between learning success and its antecedents. Second, based on our upcoming results of the objective learning success measures, we are able to provide research and practical recommendations on how the measurement of learning success should be carried out; we also derive insights on the current theoretical foundations that are usually based on self-reported measures such as dependent variables. Therefore, we hope to shed some light on the ongoing discussion of learning success measures and their validity, to finally reach consensus what learning success is, and why the measurement of it matters to research and practice (Benbunan-Fich 2010; Sitzmann et al. 2010).

Acknowledgements

Our thanks go to all participating students of the Kassel University, without whose support the study would not have been possible. We would like to thank the associate editor and the reviewers for their helpful feedback, which was extremely valuable for the present work and for future research. The research presented in this paper was funded by the German Federal Ministry of Education and Research in course of the project kultig (www.projekt-kultig.de), FKZ 01BEX05A13.

References


