



Universiteit
Leiden

The Netherlands

Transdisciplinary perspectives on validity: bridging the gap between design and implementation for technology-enhanced learning systems

Haastrecht, M.A.N. van

Citation

Haastrecht, M. A. N. van. (2025, January 24). *Transdisciplinary perspectives on validity: bridging the gap between design and implementation for technology-enhanced learning systems*. *SIKS Dissertation Series*. Retrieved from <https://hdl.handle.net/1887/4177362>

Version: Publisher's Version

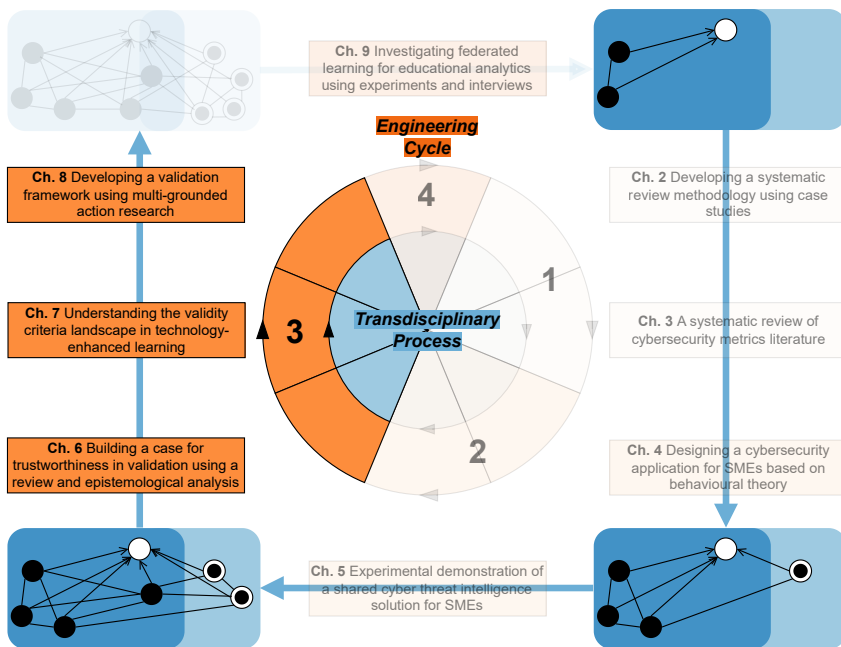
License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4177362>

Note: To cite this publication please use the final published version (if applicable).

Part III

TREATMENT VALIDATION



EMBRACING TRUSTWORTHINESS AND AUTHENTICITY IN VALIDATION

Learning analytics sits in the middle space between learning theory and data analytics. The inherent diversity of learning analytics manifests itself in an epistemology that strikes a balance between positivism and interpretivism, and knowledge that is sourced from theory and practice. In this chapter, we argue that validation approaches for learning analytics systems should be cognisant of these diverse foundations. Through a systematic review of learning analytics validation research, we find that there is currently an over-reliance on positivistic validity criteria. Researchers tend to ignore interpretivistic criteria such as trustworthiness and authenticity. In the 38 papers we analysed, researchers covered positivistic validity criteria 221 times, whereas interpretivistic criteria were mentioned 37 times. We motivate that learning analytics can only move forward with holistic validation strategies that incorporate “thick descriptions” of educational experiences. We conclude by outlining a planned validation study using argument-based validation, which we believe will yield meaningful insights by considering a diverse spectrum of validity criteria.

The contents of this chapter are based on: van Haastrecht, M. Brinkhuis, Peichl, et al. (2023). Embracing Trustworthiness and Authenticity in the Validation of Learning Analytics Systems. In Proceedings of the 13th International Learning Analytics and Knowledge Conference.

6.1 INTRODUCTION

In a recent survey among learning analytics experts (R. Ferguson et al., 2019), validity was ranked as the third-most important theme relating to the future of learning analytics, behind power (i.e., control over data) and pedagogy. R. Ferguson et al. (2019) state that validation approaches should always take “context into account when reporting results”. Recognising that each instructional context is different is seen by Gašević, Dawson, et al. (2016) as a prerequisite for an acceptable validation strategy. Kitto et al. (2018) agree, arguing that validation must address both positivistic (e.g., performance metrics) and interpretivistic (e.g., student experience) elements. They conclude that “work on developing new validation criteria that emphasise learning outcomes” is vital. This conclusion is in agreement with the experts in R. Ferguson et al. (2019), who state that “research in this space should be tied to pedagogical outcomes.”

Thus, validation is a critical topic for learning analytics research. There is agreement that validation should go beyond performance metrics and that an additional emphasis on learning outcomes would help to yield a contextualised approach. Yet, there is little consensus on which validity criteria are essential in learning analytics research. In a recent special issue on the potential links between learning analytics and educational assessment, Gašević, Greiff, et al. (2022) raised the concern that “existing learning analytic methods do not meet all of the criteria” for validation we encounter in educational assessment. However, Gašević, Greiff, et al. (2022) do not discuss to which criteria they are referring. The learning analytics literature lacks an in-depth analysis of the validity criteria that are currently in use and the criteria that deserve emphasis. We will address this gap in this chapter.

With the previous paragraphs in mind, we formulate the following main research question and sub-questions:

- **RQ:** Which validity criteria should be considered in a contextualised validation strategy for learning analytics systems?
 - **RQa:** Which validity criteria have emerged in the learning analytics domain that emphasise learning outcomes?
 - **RQb:** How has learning analytics validation research incorporated interpretivistic perspectives that recognise contextual differences?

Through an analysis of the epistemological foundations of learning analytics (Section 6.2) and a systematic review of the learning analytics validation literature (Section 6.3), we will construct an overview of emerging validity criteria to answer **RQa**. An in-depth analysis of our systematic review results (Section 6.4) will help us in answering **RQb**. We discuss the implications for our main research question in Section 6.5 and conclude in Section 6.6.

6.2 BACKGROUND: EPISTEMOLOGY AS A FOUNDATION FOR VALIDATION

How we approach validation depends on our underlying epistemology, specifically relating to our view on the concept of truth. A purely interpretivist researcher will attach little value to performance metrics when validating since they reject the concept of objective truth in social contexts. Similarly, positivist researchers are unlikely to engage in what Geertz (1973) termed “thick description” of social contexts, as they believe in the generalisability of more efficiently obtainable quantitative evidence. We posit that learning analytics epistemology is positioned in the middle space between interpretivism and positivism. In this section, we will provide further intuition for this observation and motivate that the axis of truth is not the only epistemological axis relevant to building a solid foundation for validation.

Pragmatism is one of the cornerstones of today’s learning analytics literature. As envisioned by Dewey (1931), pragmatism takes a moderate position in the interpretivism versus positivism debate. Kuhn (1962) describes the scientific process as “a process whose successive stages are characterised by an increasingly detailed and refined understanding of nature.” A process of moving “from primitive beginnings,” yet not “towards anything.” This contradicts the positivist view that the scientific method enables us to consistently hone in on truths and thereby expand our knowledge. Dewey (1938) avoids the term knowledge altogether, preferring “warranted assertability.” This phrase connects the past (warranted) and the future (assertability). Dewey’s pragmatism, therefore, blends views that aim to build from a common past (interpretivism) with those that aim to move towards a common future (positivism).

However, the axis of truth is not the only relevant epistemological axis when laying the foundations for validation. Pragmatists claim that “our conception of some given thing is bound up in our understanding of its practical application” (Knight et al., 2014). Not only a definition of what constitutes knowledge is crucial, but also a consideration of possible sources of knowledge. Pragmatism posits that practical use should be the primary source of knowledge, which juxtaposes it with rationalism which states that theoretical reasoning is the summum bonum when it comes to knowledge gathering. Wise et al. (2016) propose a similar classification regarding learning analytics design knowledge. They state that design knowledge can originate from the design process, which is guided by theory, and from the implementation process, which is coupled with the introduction of learning analytics in the learning environment.

Dewey helped develop a version of pragmatism, known as transactionalism, that emphasises contextual interactions as a vital source of knowledge (Dewey and Bentley, 1949). Transactionalism merges ideas from pragmatism and constructivism, with Dewey’s version of pragmatism being considered “as the most important precursor for social constructivism” (Reich, 2007). Social constructivism is the variant of constructivism most often encountered

in learning analytics research today. Social constructivists argue “that learners arrive at what they know mainly through participating in the social practices of a learning environment” (Woo and Reeves, 2007). Social constructivism focuses on meaningful interactions in authentic contexts. However, in today’s world, many educational interactions involve technological assistance. Although there is a role for social constructivism in technology-enhanced learning (Woo and Reeves, 2007), its focus on social interactions as the primary source of knowledge makes it ill-suited to assess the consequences of today’s socio-technical systems. Siemens (2004) aimed to solve this issue with connectivism.

Connectivism is perhaps the philosophical stance most closely associated with learning analytics. Connectivism is similar to social constructivism, but it reserves an explicit place for “learning that occurs outside of people (i.e. learning that is stored and manipulated by technology)” (Siemens, 2004). Connectivism, like learning analytics itself, states that theory is a valid source of knowledge. This brings us, finally, to the place that learning analytics epistemology occupies within the epistemological plane of Figure 6.1. We propose that learning analytics epistemology is positioned in the middle space between positivism and interpretivism on the axis of truth, but also in the middle space between theory and practice on the knowledge source axis.

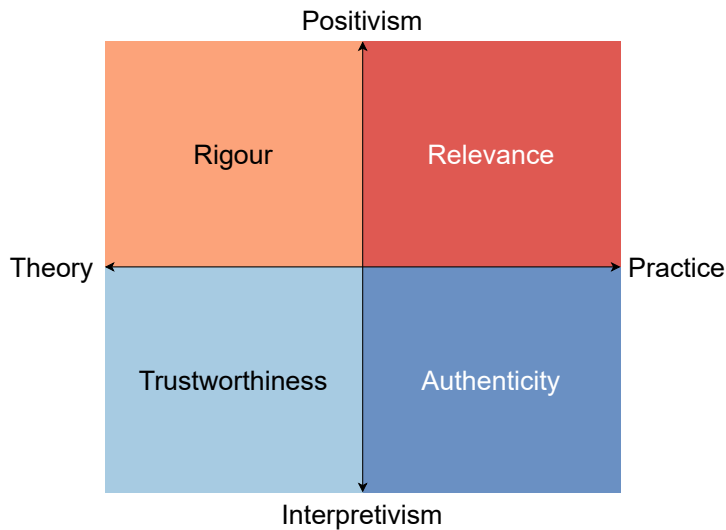


Figure 6.1: Our epistemological plane of validity, divided into four quadrants. Learning analytics occupies the middle space between positivism and interpretivism (the truth axis), and the middle space between theory and practice (the knowledge source axis).

Figure 6.1 introduces the overarching terms we use within this chapter to refer to the four quadrants created by the axes of our epistemological plane.

In the positivistic tradition, it is common to distinguish between rigour and relevance in research (Hevner et al., 2004). Rigour is connected to theory as a source of knowledge and can be “achieved by appropriately applying existing foundations and methodologies” (Hevner et al., 2004). Research is relevant when it addresses “the problems faced and the opportunities afforded by the interaction of people, organisations, and information technology” (Hevner et al., 2004). On the side of interpretivism, Guba (1981) proposed the concept of trustworthiness as a parallel to rigour. Lincoln and Guba (1986) later introduced authenticity as a more practice-oriented validity conceptualisation, noting that “conventional criteria refer only to methodology and ignore the influence of context.”

Figure 6.1 only presents four overarching validity quadrants, providing an incomplete answer to **RQa** on emerging validity criteria. Many more criteria are considered in the learning analytics literature, each with its own place within the epistemological plane. To investigate which criteria are considered and whether specific areas of the epistemological plane are underrepresented, we conducted a systematic review of the learning analytics validation literature.

6.3 METHODOLOGY

For our systematic review, we queried three databases: ACM Digital Library, Web of Science, and PubMed. We searched for all papers with abstracts containing the phrase ‘learning analytics’ and either ‘validation’ or ‘validity’. After deduplicating the query results, 83 papers remained. Of these papers, 21 formed the initial set of inclusions after excluding work that did not discuss validation or was unrelated to the field of learning analytics (as defined by SoLAR (Society for Learning Analytics Research (SoLAR), 2022)). For each of these 21 included papers, we scanned all the references and citations to find potential new inclusions. This process is known as ‘snowballing’ and is a recommended step in systematic review methodologies (van Haastrecht, Sarhan, Yigit Ozkan, et al., 2021). The snowballing phase resulted in a further 17 inclusions, meaning our final set comprised 38 papers.

Before proceeding to analyse our inclusions, we identified four papers which would allow us to construct a holistic set of potential validity criteria. We first looked towards educational measurement (sometimes referred to as educational assessment). Educational measurement is a field where validity considerations naturally take centre stage, and several learning analytics researchers have argued that we should strengthen the bond with this field (Gašević, Greiff, et al., 2022). The argument-based validation approach of Kane (2013b) has been influential in the educational measurement and learning analytics fields in recent years (Douglas et al., 2020; Milligan, 2018). Kane (2013b) stresses the importance of addressing traditional validity criteria such as rigour, construct validity, content validity, and criterion validity. However,

Kane's framework also recognises that theoretical considerations alone are insufficient, and that validation must investigate how results are used in practice. Kane captures this idea in the concept of consequential validity.

The fields of design science and information systems offer a second source of inspiration in the validity considerations made by learning analytics researchers. Mingers and Standing (2020) provide an extensive overview of the validation literature in these fields, while highlighting the importance of the interpretivistic perspective. The criteria external validity (sometimes termed generalisability), internal validity, reliability, replicability, and statistical validity occupy the rigour quadrant. Mingers and Standing (2020) additionally propose consistency (relevance quadrant) and elegance (authenticity quadrant) criteria.

Our third external source of validity terminology is the seminal interpretivistic work of Lincoln and Guba (1986). Their paper introduced the concept of authenticity as a counterbalance to trustworthiness. Lincoln and Guba (1986) discuss various dimensions of trustworthiness that parallel positivistic criteria: confirmability (related to replicability and content validity), credibility (internal validity), dependability (reliability), and transferability (external validity). They additionally discuss several dimensions of authenticity, but we select to include authenticity as a single criterion in this chapter as this is generally how the construct is viewed in learning analytics research. Lastly, Lincoln and Guba (1986) introduce fairness as a vital consideration during validation.

Finally, certain validity considerations are quite unique to the learning analytics field. To provide sufficient coverage of these validity criteria, we looked towards the work of Ali et al. (2012). They propose a diverse selection of validity criteria covering the relevance quadrant (relevance, actionability, understandability, usability, and usefulness) and the authenticity quadrant (meaningfulness and parsimony/simplicity).

6.4 RESULTS

Figure 6.2 depicts the assembled validity criteria within their respective quadrants. Criteria are positioned according to how they are defined and treated in the literature, thereby acting as a Learning Analytics Validation Assistant (LAVA). Researchers can use LAVA to determine whether the validity criteria they are considering are sufficient and appropriate for their epistemological stance. A criterion's quadrant is determined by how it is defined in one of the four core papers mentioned in the previous section. The exact placement of a criterion within a quadrant should not be interpreted as an indisputable truth. Rather, we positioned criteria relative to each other based on how they were treated and measured in the learning analytics literature.

Figure 6.2 additionally visualises the prevalence of the validity criteria in our included papers. In 38 inclusions, a total of 258 validity criteria were

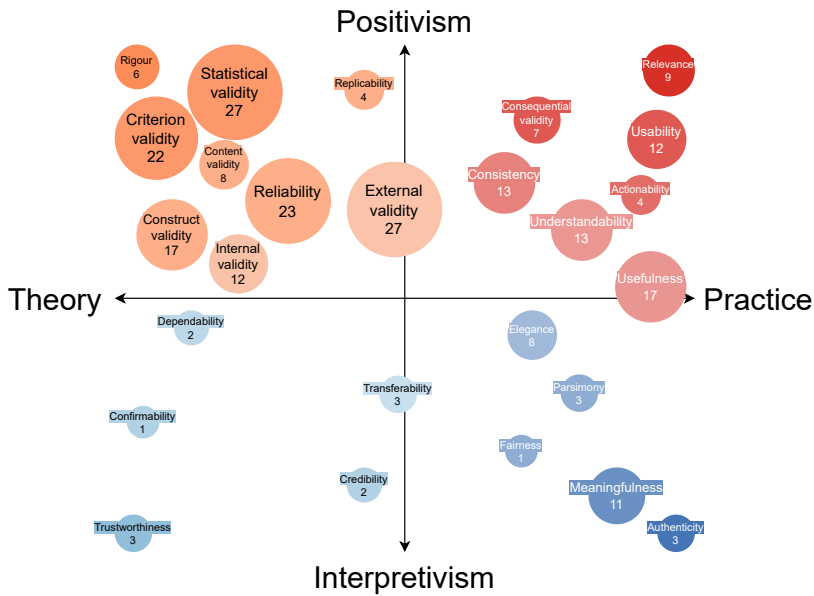


Figure 6.2: The Learning Analytics Validation Assistant (LAVA), depicting the prevalence of validity criteria observed in the learning analytics literature. Terms are positioned along the two axes (truth: positivism versus interpretivism; knowledge source: theory versus practice) of our epistemological plane.

discussed. Criteria in the rigour quadrant were mentioned 146 times (56.6%), in the relevance quadrant 75 times (29.1%), in the trustworthiness quadrant 11 times (4.3%), and in the authenticity quadrant 26 times (10.1%). Hence, researchers covered positivistic criteria 221 times, whereas interpretivistic criteria were mentioned only 37 times.

Statistical validity and external validity are the criteria mentioned most often within our inclusions. For statistical validity, we noticed that most papers focus on statistical significance, whereas Saqr and López-Pernas (2021) point out that researchers should additionally consider effect size. External validity is another problematic criterion within learning analytics research. Of the 27 times external validity was mentioned in one of our inclusions, 24 times the authors concluded that the external validity of their study was lacking. We observed a similar pattern with the interpretivistic counterpart to external validity: transferability. Of the three times transferability was considered, the authors stated on two occasions that more work was necessary to assess the transferability of their results.

Figure 6.2 provides an answer to **RQa**: Which validity criteria have emerged in the learning analytics domain that emphasise learning outcomes? Criteria on the ‘practice’ half of the diagram relate to outcomes of the learning process. Criteria positioned on the extreme right of the theory-practice axis correspond to an advanced internalisation of learning analytics outcomes. Learning analytics researchers evidently attach importance to relevant, usable, actionable, and useful solutions. Additionally, several papers recognised that authentic, meaningful learning experiences are not simply a luxury, but a goal to strive for.

Table 6.1 lists the combinations of validity criteria quadrants observed in our inclusions. Four out of 38 papers covered criteria from all four quadrants. Papers tended to consider criteria from at least two quadrants, with only one inclusion not covering a criterion from the rigour quadrant. Conversely, the criteria in the trustworthiness quadrant, along with parsimony, authenticity, and fairness, are mentioned least often. Many of these criteria can only be assessed through “thick descriptions” of social contexts (Geertz, 1973), possibly pointing to barriers to engaging in such activities within learning analytics research. Moreover, although meaningfulness was discussed in 11 papers, only one of these papers conducted qualitative interviews during validation. All other papers used either quantitative data analysis or structured questionnaires in their evaluation. Concerning **RQb**, we can conclude that although interpretivistic validity criteria are considered in learning analytics research, their treatment is often too superficial to provide in-depth insight into contextual learning experiences.

Table 6.1: Combinations of the four validity criteria quadrants (rigour, relevance, trustworthiness, and authenticity) observed in the 38 inclusions of our systematic review, sorted by number of related inclusions. Only observed combinations are listed.

QUADRANT COMBINATION	RELATED INCLUSIONS
Rigour, relevance	Berman and Artino (2018), Chaparro-Peláez et al. (2020), Dourado et al. (2021), Effenberger and Pelánek (2021), Fincham et al. (2019), Galaige et al. (2018), Giannakos et al. (2015), Howell et al. (2018), Saqr, Viberg, et al. (2020), and Tabuenca et al. (2015)
Rigour, relevance, authenticity	Alonso-Fernández et al. (2019), Ifenthaler and Widanapathirana (2014), Kärner et al. (2021), Muñoz et al. (2020), Pardo et al. (2015), Park and Jo (2019), Saqr and López-Pernas (2021), Whitelock-Wainwright et al. (2020), Ye and Pennisi (2022), and Zheng et al. (2022)
Rigour	Bitner et al. (2020), Jo et al. (2014), Maldonado-Mahauad et al. (2018), Matcha et al. (2020), and Prat and Code (2021)
Rigour, authenticity	Chejara et al. (2021), Fan, Lim, et al. (2022), D. Kim et al. (2016), Kizilcec et al. (2017), and Sinha et al. (2014)
Rigour, relevance, trustworthiness, authenticity	Ali et al. (2012), Valle et al. (2021), Wise et al. (2016), and Yoo and Jin (2020)
Rigour, relevance, trustworthiness	Cerro Martínez et al. (2020), Saqr, Fors, et al. (2018), and Winne (2020)
Relevance	Gañán et al. (2017)

6.5 DISCUSSION

Our results lead to three main findings related to the learning analytics validation literature, which we will cover in this section.

6.5.1 *Troubling external validity*

Learning analytics researchers seem to have a troubling relationship with external validity. Together with statistical validity, external validity was the criterion mentioned most often in our inclusions. Yet, 24 out of the 27 papers that mention external validity conclude that there are limitations to the generalisability of their results. At times, the limited scale of studies is listed as the cause for generalisability concerns (e.g., (Chaparro-Peláez et al., 2020; Tabuenca et al., 2015; Wise et al., 2016; Yoo and Jin, 2020)). Elsewhere, researchers provide a general warning that more research is necessary should one want to generalise the results (e.g., (Effenberger and Pelánek, 2021; D. Kim et al., 2016; Saqr, Fors, et al., 2018)). Transferability, the interpretivistic parallel of external validity, suffers from the same issue. Researchers state that results could be transferred to other contexts, but that more research is required to confirm this claim (Ali et al., 2012; Cerro Martínez et al., 2020).

The reader should not interpret the previous paragraph as a critique of the cited research. If there are limitations to the generalisability of findings, these should be mentioned. However, we should avoid a situation in the learning analytics field where generalisability becomes an afterthought that can always be left for future work. External validity and transferability are valued validity criteria that should guide learning analytics research *a priori*, not *a posteriori*.

Replication studies that aim to understand the validity of learning analytics solutions in new contexts should receive more attention.

6.5.2 *A need for thick descriptions*

We noted in Section 6.4 that even papers that recognise interpretive validity criteria (e.g., meaningfulness) often resort to quantitative methods during validation. Geertz (1973) believes that the analysis of social culture and context requires qualitative methods “in search of meaning” rather than quantitative methods “in search of law.” In other words, we require “thick descriptions” of the educational contexts being considered in learning analytics research. Thick descriptions that cannot be obtained through data analysis or questionnaires, but that require qualitative methods.

The advantages of using qualitative methods go beyond a deeper understanding of the educational context. As Guba (1981) recognises, “to determine the extent to which transferability is probable, one needs to know a great deal about both the transferring and receiving contexts.” Guba (1981) states that thick descriptions are essential if we wish to achieve transferable results. Thus, thick descriptions provide deeper insight into interpretivistic validity criteria and concurrently act as a catalyst in facilitating generalisable learning analytics research. Researchers looking to produce more generalisable results will benefit from employing qualitative research methods such as qualitative interviews and action research.

6.5.3 *The potential of argument-based validation*

To conclude this section, we will discuss a validation approach uniquely suited to facilitate the diverse validity criteria and research methods covered in this chapter: argument-based validation. Kane (2013b) originally introduced this approach in the educational measurement field. Gašević, Greiff, et al. (2022) argue that learning analytics research can profit from the vast validity experience within educational measurement and psychological assessment, and argument-based validation has started to see use within the learning analytics domain (Douglas et al., 2020; Milligan, 2018).

In general, research uses inferences to make warranted claims based on data. Argument-based validation proceeds by constructing arguments to provide evidence for the assertability of these claims. Once evidence has been assembled in structured arguments, we assess the validity of the overall inference chain. The benefit of this approach is that it gives a balance of flexibility and structure, allowing researchers to recognise “legitimately diverse arguments” (Addey et al., 2020) while avoiding the open-ended nature of validation. The original framework of Kane (2013b) has been extended to allow for an increased focus on practical consequences (Hopster-den Otter et al., 2019) and to

explicitly address fairness in artificial intelligence (AI) enhanced assessments (Huggins-Manley et al., 2022). Argument-based validation is a promising avenue for learning analytics researchers looking to address diverse validity criteria and produce rigorous, relevant, trustworthy, and authentic results.

6.6 CONCLUSION AND FUTURE WORK

Within this chapter, we have investigated which validity criteria should be considered in a contextualised validation strategy for learning analytics systems. We proceeded by first analysing the epistemological foundations of learning analytics research, concluding that learning analytics epistemology is positioned in the middle space between positivism and interpretivism and between theory and practice. We then conducted a systematic review to uncover which types of validity criteria are employed by learning analytics researchers. We visualised the results to create a Learning Analytics Validation Assistant (LAVA).

We uncovered an over-reliance on positivistic criteria. Interpretivistic criteria that were covered (e.g., meaningfulness), were often investigated using quantitative rather than qualitative methods. In Section 6.5, we analysed the LAVA results and delineated a need for more focus on “thick descriptions” of educational experiences. Such thick descriptions help to foster a deeper understanding of the context being studied and can act as a catalyst in facilitating generalisable research.

In future work, we will apply our LAVA insights within an educational research project. As suggested in Section 6.5.1, we intend to employ an argument-based validation approach incorporating diverse arguments and validity criteria. We recognise that we are bound to encounter limitations in our future work and want to stress that no single approach can function as a validation panacea. Nevertheless, we believe that LAVA can stimulate researchers to evaluate whether their validity criteria are sufficient and appropriate for their epistemological stance.