Empowering higher education students to monitor their learning progress: opportunities of computerised classification testing

Dirk Ifenthaler

Learning, Design and Technology, University of Mannheim, Mannheim, Germany and UNESCO Deptuy-Chair on Data Science in Higher Education Learning and Teaching, Curtin University, Perth, Australia, and

Muhittin ŞAHİN

Learning, Design and Technology, University of Mannheim, Mannheim, Germany

Abstract

Purpose – This study aims to focus on providing a computerized classification testing (CCT) system that can easily be embedded as a self-assessment feature into the existing legacy environment of a higher education institution, empowering students with self-assessments to monitor their learning progress and following strict data protection regulations. The purpose of this study is to investigate the use of two different versions (without dashboard vs with dashboard) of the CCT system during the course of a semester; to examine changes in the intended use and perceived usefulness of two different versions (without dashboard vs with dashboard) of the CCT system; and to compare the self-reported confidence levels of two different versions (without dashboard vs with dashboard) of the CCT system.

Design/methodology/approach – A total of $N = 194$ students from a higher education institution in the area of economic and business education participated in the study. The participants were provided access to the CCT system as an opportunity to self-assess their domain knowledge in five areas throughout the semester. An algorithm was implemented to classify learners into master and nonmaster. A total of nine metrics were implemented for classifying the performance of learners. Instruments for collecting co-variates included the study interest questionnaire (Cronbach's $a = 0.90$), the achievement motivation inventory (Cronbach's $a = 0.94$), measures focusing on perceived usefulness and demographic data.

Findings – The findings indicate that the students used the CCT system intensively throughout the semester. Students in a cohort with a dashboard available interacted more with the CCT system than students in a cohort without a dashboard. Further, findings showed that students with a dashboard available reported significantly higher confidence levels in the CCT system than participants without a dashboard.

Originality/value – The design of digitally supported learning environments requires valid formative (self-)assessment data to better support the current needs of the learner. While the findings of the current study are limited concerning one study cohort and a limited number of self-assessment areas, the CCT system is being further developed for seamless integration of self-assessment and related feedback to further reveal unforeseen opportunities for future student cohorts.

Keywords Computerized adaptive assessment, Computer-based assessment, Self-assessment, Data analytics, Classification, Feedback, Higher education, University

Paper type Research paper

1. Introduction
Digitally supported assessment systems provide opportunities for supporting learning processes and learning outcomes (Pachler *et al.*, 2010). To facilitate learning through assessment, Carless (2007) emphasizes that assessment tasks should be learning tasks that are related to the defined learning outcomes and distributed across the learning and course period. Furthermore, to foster learners' responsibility for learning (Bennett, 2011; Wanner and Palmer, 2018) and self-regulation (Panadero *et al.*, 2017), self-assessments are suitable means. Sadler (1989) argues that self-monitoring and external feedback are related to formative assessment, with the aim of evolving from using external feedback to self-monitoring to independently identify gaps for improvement. Hence, self-assessments enable learners to develop independence from relying on external feedback (Andrade, 2010). However, making use of self-assessments might be particularly challenging for learners with lower levels of domain or procedural knowledge (Sitzmann *et al.*, 2010).

Terms relating to digitally supported assessment systems have been used inconsistently. Frequent terms include computer-based assessment and computer-based testing (Quellmalz, 2015), computerized mastery testing (Liefeld and Herrmann, 1990) or computer-administered testing (Carlson, 2015). Given the availability of advanced data analytics applications, computerized adaptive testing (CAT) and computerized classification testing (CCT) systems have seen a rise in implementation (van der Kleij and Adie, 2018). These systems aim at classifying learners using two or more categories rather than determining their ability estimate (Lin and Spray, 2000). In addition, item pools of the CAT need not be as large as approaches of classic testing (Parshall *et al.*, 2002). However, as noted earlier by Ellis (2013), data analytics approaches still fail to make full use of educational technology and data for assessment.

While advances in research on online self-assessments and related systems are growing rapidly (Heil and Ifenthaler, 2023), higher education institutions lack organization-wide implementation of sustainable technology innovation (Buckingham Shum and McKay, 2018; Ifenthaler, 2017). Therefore, this project thought to implement and further advance a CCT system for online self-assessment in a productive learning environment of a higher education institution, examine students' usage behaviors of the CCT system and determine students' intended use, perceived usefulness as well as their self-reported confidence levels after using CCT system.

2. Literature review
2.1 Online self-assessment in higher education
Online self-assessment in higher education describes the assessment of students learning with digital tools, including information and communication technologies (Conrad and Openo, 2018). This does not restrict online self-assessment to fully online courses and can also be

implemented in a blended learning format (Gikandi *et al.*, 2011). Online self-assessments may take on different pedagogical functions as part of online learning environments (Webb and Ifenthaler, 2018), for example, scaffolding students to complete a task and measuring how much support they need (Ahmed and Pollitt, 2010) or providing students with semantically rich and personalised feedback, as well as adaptive prompts for reflection (Ifenthaler, 2012; Schumacher and Ifenthaler, 2021). Other examples of online self-assessments include a pedagogical agent acting like a virtual coach tutoring learners and providing feedback when needed (Johnson and Lester, 2016) as well as an analysis of a learner's decisions during a digital game or simulation (Bellotti *et al.*, 2013). Other online self-assessments use multimedia-constructed response items for authentic learning experiences (Lenhart, 2015) or provide students with an emotionally engaging virtual world experience that unobtrusively documents the progression of a person's leadership and ethical development over time (Turkay and Tirthali, 2010). Thus, online self-assessments offer a broad range of pedagogical functions, including a medium for communication, a learning assistant, a judge, a test administrator, a performance prompt, a practice arena or a performance workspace (Webb *et al.*, 2013).

To facilitate learning, online self-assessment tasks should be learning tasks that are related to the defined learning outcomes and distributed across the learning and course period (Carless, 2007). Furthermore, online self-assessments are a useful method for encouraging learners' ownership (Bennett, 2011; Wanner and Palmer, 2018) of their learning and self-regulation (Panadero *et al.*, 2017). Hence, online self-assessments enable learners to develop independence from relying on external feedback (Andrade, 2010). However, online self-assessments demand and foster evaluative judgment of learners (Tai *et al.*, 2018). Thus, online self-assessments might be particularly challenging for learners with lower levels of domain or procedural knowledge (Sitzmann *et al.*, 2010). Hence, the feedback generated internally by the learners could be complemented and further enhanced with external feedback (Butler and Winne, 1995). Such external feedback may help learners adjust their self-monitoring (Sitzmann *et al.*, 2010). Among others, the feedback provided should clearly define expectations (i.e. criteria, standards, goals), be timely, sufficiently frequent and detailed, be on aspects that are malleable through the students, be on how to close the gap, in a way learners can react upon it (Nicol and Macfarlane-Dick, 2006).

It is expected that digitally supported and data-driven systems may enhance external feedback while meeting several specific requirements, such as follows:

- adaptability to different subject domains;
- flexibility for experimental as well as learning and teaching settings;
- management of huge amounts of data;
- rapid analysis of complex and unstructured data;
- immediate feedback for learners and educators; and
- generation of automated reports of results for educational decision-making (Ifenthaler *et al.*, 2010).

2.2 Computerized classification testing
CCT has a long history in psychological and educational research as well as pedagogical practice (van der Linden and Glas, 2000). CCT systems aim at classifying learners into two or more categories (van Groen, 2012). Classifications with two categories (Huebner, 2012; van Groen *et al.*, 2019), as well as three and more categories (Eggen and Straetmans, 2000), are frequently implemented. CCT uses various methodological approaches and algorithms

for classifying learners with the least number of items (Thompson, 2007). The Sequential Probability Ratio Test (SPRT) is a frequently used algorithm in CCT systems. SPRT follows a decision matrix to decide which one out of two hypotheses is more correct (Wald, 1947). Such algorithms enable CCT systems to select and present the most appropriate assessment items to individual learners (Spray and Reckase, 1996) in comparison to an expected standard or predefined benchmark (Parshall *et al.*, 2002). According to Frick (1990), SPRT algorithms are less complex and more practical for implementation as well as require less time for rendering decisions. For instance, Frick (1992) found that the SPRT algorithm could classify a learner as a master (advanced learner in a specific knowledge domain) vs a nonmaster (novice learner in a specific knowledge domain) using an average of ten assessment items.

2.3 Data-driven dashboards
Dashboards are customizable control panels displaying features that may adapt to a specified process in real-time. Dashboards in the context of learning analytics are being developed and implemented to visualize the analytics results of learner-generated data and other relevant information. Such visualizations are expected to create awareness and reflection among learners (Roberts *et al.*, 2017). The functions of visualizations include exploration, discovery, summarising, presenting, comparing and enjoying (Verbert *et al.*, 2014).

Current research on dashboards aims to identify which data are meaningful to different stakeholders in education and how data can be presented to support learning processes and outcomes (Bodily and Verbert, 2017; Sahin and Ifenthaler, 2021). The objectives of recent dashboard research include the following:

- increasing awareness about the learning process;
- supporting cognitive processes;
- identifying students at risk;
- providing immediate feedback;
- displaying achievement level;
- providing procedural information;
- supporting decision-making;
- informing;
- showing participant relationships;;
- comparing with peers; and
- reflecting learning activities.

Most visualization techniques stem from statistics, including bar charts, line graphs, tables, pie charts and network graphs.

A recent systematic literature review identified 76 studies focusing on various dashboard features and stakeholder applications (Sahin and Ifenthaler, 2021). Contrary to previous findings (Leitner *et al.*, 2019), this current state of research identified an increase in empirical research studies. Nevertheless, the available research does not include sufficient experimental and controlled evidence with a specific focus on the design of dashboards for online self-assessment.

2.4 Current study
This project focused on providing a CCT system that can easily be embedded as an online self-assessment feature into the existing legacy environment of higher education

institutions, empowering students with online self-assessments to monitor their learning progress and following strict data protection regulations.

This study has three aims:

(1) to investigate the interaction with two different versions of the CCT system (without dashboard vs with dashboard) during the course of a semester;

(2) to examine changes in the intended use and perceived usefulness of two different versions (without dashboard vs with dashboard) of the CCT system; and

(3) to compare the self-reported confidence levels of two different versions (without dashboard vs with dashboard) of the CCT system.

3. Method

3.1 Design and participants

Through evidence-based design, the implementation of the CCT system for supporting online self-assessments was part of a larger initiative for establishing data-driven features into the existing legacy system of the higher education institution. For instance, learning analytics features and promoting functions were implemented into the learning management system (Klasen and Ifenthaler, 2019).

This design-based research study was conducted over the course of two semesters for two similar study cohorts as part of a lecture in a bachelor's program on research methodology in the field of education. In the first iteration of the CCT system, students had the opportunity to interact with the CCT system without the support of a dashboard. The second iteration of the CCT system included a dashboard focussing on individual learning progression (Figure 1) and group comparisons (Figure 2).
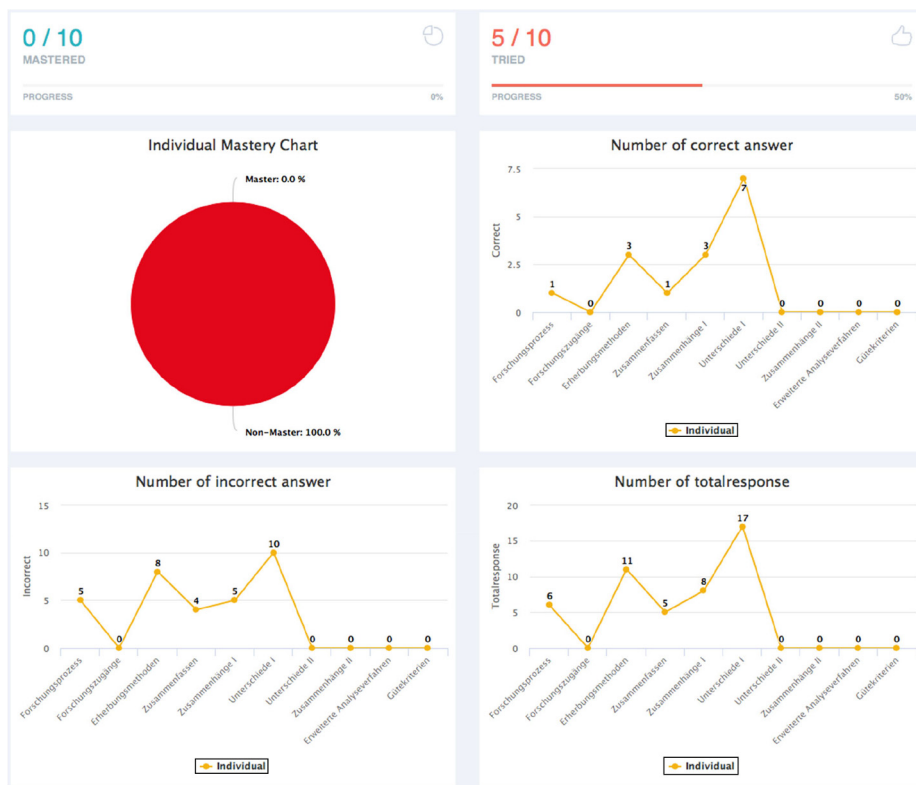
The dashboard included information on the individual self-assessment results, i.e. the number of mastery subject areas, number of correct answers, number of incorrect answers, number of total responses and number of total attempts. In addition to the information displayed on the individual dashboard (Figure 1), the group dashboard included averaged information about other anonymized learners, which enabled the learner to compare individual and group performance (Figure 2).

A total of $N = 194$ students (138 female; 56 male) from a European university in the area of economic and business education participated in the study. The participants' average age was 23.15 years ($SD = 2.34$), while they studied for an average of 5.71 semesters ($SD = 1.49$). The first study cohort (without a dashboard; SC1) included $N = 107$ students (71 female; 36 male), and their average age was 23.32 ($SD = 2.51$). The second cohort (with dashboard; SC2) consisted of $N = 87$ students (67 female; 20 male) with an average age of 22.94 ($SD = 2.11$). Both study cohorts are similar with regard to their course enrolment, prior knowledge and study experience. Ethics consent was obtained for this research project.

3.2 Computerized classification testing system

At the start of the semester, the CCT system was embedded as an online self-assessment feature in the productive learning management system. Various subject areas for self-assessment were defined:

- research approaches;
- research process;
- research designs;
- statistical correlations;

Figure 1.
CCT dashboard
including individual
learning progression

Source: Figure by authors

- advanced statistical correlations;
- significant differences;
- advanced significant difference; and
- research quality criteria.

A total of $N = 256$ assessment items (true/false and multiple-choice) were available in the item bank. The CCT system was available throughout the semester.

The CCT SPRT algorithm was implemented for classifying learners into two categories (van Groen, 2012): master and nonmaster. In addition, a random item selection feature was implemented. Figure 3 shows an example of an individualized performance chart learners received after a completed round of online self-assessments. The charts provide an overview of the individual task performance (blue line in Figure 3) as well as the thresholds for being master (red line in Figure 3) and nonmaster (yellow line in Figure 3).

A total of nine indicators were implemented for classifying the performance of learners (see Table 1). Principal component analysis (PCA) as a feature selection
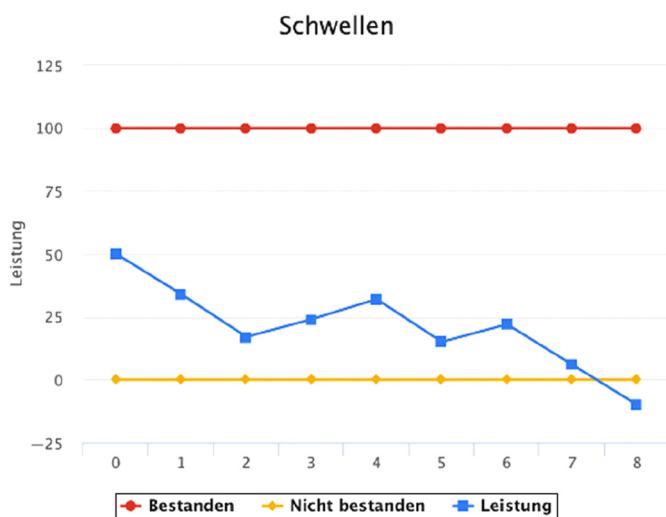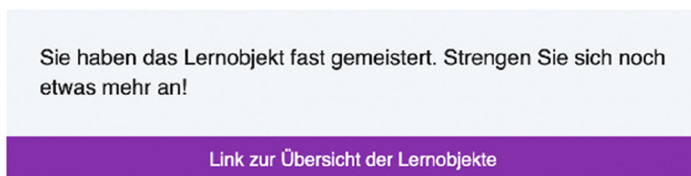
Figure 2.
CCT dashboard
including group
comparison

Source: Figure by authors

algorithm was conducted to determine which of the indicators collected by the CCT system are more important in students' behavioral engagement (Table 1). PCA is used to reduce the number of components in the available data set (Fabrigar *et al.*, 1999). Accordingly, it is expected that the communality values of the indicators are high. Factor scores were examined to determine which indicators provided more information, i.e. the importance of participants' engagement with the CCT system. The most important indicators for learners' behavioral engagement in the CCT system are the number of responses, the number of attempts, the number of correct answers and the number of incorrect answers. These indicators explain 91% of learners' behavioral engagement structure. These indicators were implemented in the CCT system dashboards.

3.3 Materials and instruments
Instruments for collecting co-variates included the FSI (Schiefele *et al.*, 1993), a study interest questionnaire (Cronbach's α = 0.90), the short version of the LMI-K (Schuler and Prochaska, 2001), which is an achievement motivation inventory (Cronbach's α = 0.94), measures

Figure 3.
CCT performance
chart after
completing online
self-assessments

Source: Figure by authors

CCT system indicators	Communality	Factor score
Number of responses	0.992	0.996
Test attempts	0.958	0.979
Number of correct answer	0.931	0.965
Number of incorrect answer	0.767	0.876
Time spent/login duration	0.560	0.748
Number of master subject/areas	0.893	0.680
Number of subject attempt	0.808	0.676
Login frequency	0.503	0.655
Number of nonmaster subject/areas	0.360	0.404

Source: Table by authors

Table 1.
Communality and
factor score of the
CCT system
variables

focusing on perceived usefulness (Cronbach's α = 0.90) (Davis, 1989) and confidence (Cronbach's α = 0.88) as well as demographic data (e.g. gender, age, study experience).

FSI and LMI-K data collection tools were used for internal validity. Learners' confidence and perceived usefulness level were tested according to the study cohorts (SC1: CCT without dashboard; SC2: CCT with dashboard).

3.4 Procedure and analysis
Participants received a brief introduction to the CCT system and could access it anytime during the semester. At the end of the semester, participants were asked to complete the following surveys:

- FSI;
- LMI-K;
- perceived usefulness;
- confidence in the CCT system; and
- demographic data.

The course performance was assessed through an open-ended exam with a duration of 90 min. The examined questions related to the five areas of the CCT system. The exam scores were classified into high and low performance, where low performance indicated an increased need for learning support.

Log data from the CCT environment were collected using a time-stamped sequence format. Log data of the CCT system represented two categories:

(1) interactions with the assessment items; and
(2) interactions with the dashboard.

Interactions with the assessment items included the login frequency, number of correct responses, number of master subjects, number of incorrect answers and number of total attempts. The dashboard interactions considered the time spent on individual or group dashboards.

As a standard data-protection practice, all data were stored and analyzed using an anonymized procedure. Time-stamped log-data from the CCT system included the following:

- login frequency;
- login duration;
- correct answers;
- incorrect answers;
- number of responses;
- master subject areas;
- nonmaster subject areas;
- test attempts; and
- subject area attempts.

Data were cleaned and combined for descriptive and inferential statistics using SPSS version 27.

4. Results
4.1 Computerized classification testing system interaction
For answering the first research question, the participants' CCT system interactions were determined based on the CCT system's log-data. Log-data consisted of the number of responses, test attempts, number of correct answers and number of incorrect answers. We computed an independent *t*-test to check for differences in participants' CCT system interaction between the study cohorts (SC1 vs *SC2*). The independent *t*-test analysis

revealed a highly significant difference in CCT system interaction between SC1 (without dashboard; $M = -0.31$; $SD = 0.59$) and SC2 (with dashboard; $M = 0.37$; $SD = 1.25$), $t(192) = 4.936$, $p < 0.001$, $d = 0.69$.

Accordingly, participants of the second study cohort (SC2; dashboard available) interacted significantly higher with the CCT system than participants of the first study cohort (SC1; no dashboard).

4.2 Intended use and perceived usefulness

Regarding the second research question, the participants self-reported intended use and perceived usefulness were analyzed. We computed two independent t-tests to check for differences in participants' intended use and perceived usefulness of the CCT system between the study cohorts (SC1 vs SC2). The first independent t-test analysis revealed a significant difference in intended use between SC1 (without dashboard; $M = 17.72$; $SD = 2.85$) and SC2 (with dashboard; $M = 16.59$; $SD = 3.11$), $t(192) = 2.641$, $p < 0.01$, $d = 0.38$. The second independent t-test analysis revealed a significant difference of perceived usefulness between SC1 (without dashboard; $M = 20.87$; $SD = 3.55$) and SC2 (with dashboard; $M = 19.47$; $SD = 3.72$), $t(192) = 2.672$, $p < 0.01$, $d = 0.39$.

Accordingly, participants of the first study cohort (SC1; without dashboard) reported significantly higher intended use and perceived usefulness of the CCT system than participants of the second study cohort (SC2; with dashboard).

4.3 Confidence

For answering the third research question, the participant's self-reported confidence level was examined. We computed an independent t-test to check for differences in participants' confidence level regarding the CCT system between the study cohorts (SC1 vs SC2). The independent t-test analysis revealed a significant difference in the confidence level between SC2 (with dashboard; $M = 7.71$; $SD = 1.81$) and SC1 (without dashboard; $M = 7.08$; $SD = 1.80$), $t(192) = 2.415$, $p < 0.01$, $d = 0.35$.

Accordingly, participants of the second study cohort (SC2; dashboard available) reported significantly higher confidence level in the CCT system than participants of the first study cohort (SC1; no dashboard).

5. Discussion

The complexity of designing technology- and analytic-enhanced assessment and feedback systems has been discussed widely over the past few years (Sadler, 2010; Shute, 2008; Webb and Ifenthaler, 2018). Online assessment may be implemented on platforms such as Learning Management Systems, through game-based environments, or on specific websites or applications (e.g. ePortfolios). However, an online assessment might lead to an increase in academic misconduct through unsupervised use of the systems (Tsai, 2016). Analytics-enhanced assessment systems may be used to detect academic dishonesty, but similarly, when implementing these types of approaches, practitioners should consider ethical and privacy issues and ensure not creating a feeling of surveillance (Gašević *et al.*, 2022).

This project aimed to implement and advance a CCT system for online self-assessment in a productive learning environment of a higher education institution, examine students' CCT usage behaviors and ascertain students' intended use, perceived usefulness and their self-reported confidence levels after using the CCT system.

5.1 Summary of findings

The findings indicate that the students used the CCT system intensively throughout the semester. In-depth PCA revealed that the number of total responses, the number of attempts, correct answers and incorrect answers seem to be valid metrics for being included in the design of future self-assessment or learning analytics systems (Park and Jo, 2015; Tempelaar *et al.*, 2015).

The first research question investigated the use of the two different versions (without dashboard vs with dashboard) of the CCT system during the course of a semester. The findings revealed a significant difference between the two study cohorts, with a medium effect of more interactions by students having the dashboard available. The more options a digital system provides, the more interactions are expected. In-depth post-hoc log-data analysis confirmed that students spent similar time with the self-assessment items; however, additionally used the dashboard for reflecting their learning progress (Roberts *et al.*, 2017).

The second research question examined the intended use and perceived usefulness of two different versions (without dashboard vs with dashboard) of the CCT system. Contrary to the expectations based on previous research, students of the first study cohort (SC1; without dashboard) reported significantly higher intended use and perceived usefulness of the CCT system than participants of the second study cohort (SC2; with dashboard) (both differences included a small effect). While the dashboard design followed current state-of-the-art visualization features (e.g. line graphs) (Schwendimann *et al.*, 2016), students were unable to make use of the visualization to foster their learning process. Accordingly, dashboards may require features students want for supporting their learning, e.g. personalised scaffolds or adaptive content recommendations (Schumacher and Ifenthaler, 2018).

The third research question tested for differences concerning the confidence level toward the CCT system. Interestingly, students of the second study cohort (SC2; dashboard available) reported significantly higher confidence levels of the CCT system than participants of the first study cohort (SC1; no dashboard) (small effect). In contrast to the intended use and perceived usefulness of the CCT system, the available dashboard seems to increase confidence in using the CCT system. Hence, data visualisations related to self-assessments seem to foster students' trust in CCT systems (Pardo and Siemens, 2014).

Additional analysis focusing on differences in the psycho-educational structures of the students were tested. FSI and LMI did not reveal a statistically significant difference between the two study cohorts.

5.2 Research implications and contribution

The findings support the assumption that self-assessments and related feedback support learning processes and impact the learning performance (Adler and Benbunan-Fich, 2015; Azevedo and Bernard, 1995). Thus, students who were classified with more master subject areas outperformed students who were classified with fewer master subject areas. However, in contrast to previous findings (Jo *et al.*, 2015), students' characteristics did not contribute to the prediction of the student's performance.

The reported small and medium effect sizes highlight the importance of further understanding what learners expect from learning analytics dashboards and related visualisations (Bennett and Folley, 2021; Schumacher and Ifenthaler, 2018). Visualisation techniques such as line chart, bar chart, progress bar, timeline and pie chart seem to be limited in supporting student learning (Sahin and Ifenthaler, 2021). Accordingly, inappropriate designs can negatively affect the learning processes of the learners (Bodily *et al.*, 2018; Schwendimann *et al.*, 2016). To design an effective dashboard design, it is necessary to

establish a theoretical connection with human cognition and perception, situation awareness and visualisation technologies (Yoo *et al.*, 2015). In addition, dashboard visualisations may include contextually appropriate presentations, visual language and social framing (Sarikaya *et al.*, 2018). Further, affective dispositions of learners which may be triggered through dashboards and visualisations could be considered when designing future implementations, for instance, understanding the need for autonomy, relatedness or interest when designing learning analytics dashboards (Eseryel *et al.*, 2014; Howell *et al.*, 2018; Roberts *et al.*, 2017). Hence, the acceptance and effectiveness of dashboards highly depend on the benefits learners may expect, including a clear, simple and fit-for-purpose design (Pokhrel and Awasthi, 2021). However, dashboard designs are expected to be unique regarding the design of the visualisation and are also expected to fit the current needs of learners when they are used (Teasley *et al.*, 2021). While the currently implemented dashboard is being further developed, the learners expectations are being evaluated in a participatory design approach (Könings *et al.*, 2014).

5.3 Limitations and future research

This study has several limitations that need to be addressed. The sample included a select group of participants from one university all enrolled in a specific course, thus prohibiting generalisations of results. This fact limits the external validity of our findings (Campbell and Stanley, 1963). Accordingly, future studies shall include participants within and across different subject domains and from different higher education institutions. The administered self-reporting inventories for assessing participants' dispositions are limited as they can only gather perceptions after using the CCT system and of which the learners are aware (Veenman, 2013). Still, using other methodological approaches to investigate dispositions toward the CCT system, such as think-aloud protocols influence the behavior as students might be more aware of their actions or might feel interrupted (Schraw, 2010) (Schraw, 2010). Therefore, opportunities for a combination of a multi-method approach to assess learners' dispositions seem to be a reasonable future approach (Azevedo *et al.*, 2010). Last, while our sample was large enough to achieve statistically significant results, the explained variance and respected effect sizes were rather moderate. This indicates that besides the tested variables others may have influenced the outcomes that were not tested in this study.

Accordingly, the CCT system is being further developed for seamless integration of self-assessment and related feedback to further reveal unforeseen opportunities for future student cohorts (Saqr *et al.*, 2017). As the next step of this research project, dynamic student-facing dashboards will be implemented that are based on assessment indicators (Bodily *et al.*, 2018). Different dashboard designs will be tested in quasi-experimental studies that will be embedded in a productive digital learning environment. It is thought that these dashboards provide students with meaningful insights into monitoring their performance and their deficiencies or strengths. Thus, it will be possible for learners to control their learning processes and support learners' autonomy (Carless and Boud, 2018; Matcha *et al.*, 2020). In conclusion, the design of digitally supported learning environments requires valid formative (self-assessment) data to better support the current needs of the learner.

References

Adler, R.F. and Benbunan-Fich, R. (2015), "The effects of task difficulty and multitasking on performance", *Interacting with Computers*, Vol. 27 No. 4, pp. 430-439.

Ahmed, A. and Pollitt, A. (2010), "The support model for interactive assessment", *Assessment in Education: Principles, Policy and Practice*, Vol. 17 No. 2, pp. 133-167.

Andrade, H.J. (2010), "Students as the definitive source of formative assessment: academic self-assessment and the self-regulation of learning", in Andrade, H.J. and Cizek, G.J. (Eds), *Handbook of Formative Assessment*, Routledge, New York, pp. 90-105.

Azevedo, R. and Bernard, R.M. (1995), "A meta-analysis of the effects of feedback in computer-based instruction", *Journal of Educational Computing Research*, Vol. 13 No. 2, pp. 111-127.

Azevedo, R., Johnson, A., Chauncey, A. and Burkett, C. (2010), "Self-regulated learning with MetaTutor: advancing the science of learning with metacognitive tools", in Khine, M.S. and Saleh, I.M. (Eds), *New Science of Learning*, Springer, New York, pp. 225-247.

Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P. and Berta, R. (2013), "Assessment in and of serious games: an overview", *Advances in Human-Computer Interaction*, Vol. 2013, p. 136864.

Bennett, R.E. (2011), "Formative assessment: a critical review", *Assessment in Education: Principles, Policy and Practice*, Vol. 18 No. 1, pp. 5-25, doi: 10.1080/0969594X.2010.513678.

Bennett, L. and Folley, S. (2021), "Students' emotional reactions to social comparison via a learner dashboard", in Sahin, M. and Ifenthaler, D. (Eds), *Visualizations and Dashboards for Learning Analytics*, Springer, Cham, pp. 233-249, doi: 10.1007/978-3-030-81222-5_11.

Bodily, R. and Verbert, K. (2017), "Review of research on student-facing learning analytics dashboards and educational recommender systems", *IEEE Transactions on Learning Technologies*, Vol. 10 No. 4, pp. 405-418.

Bodily, R., Ikahihifo, T.K., Mackley, B. and Graham, C.R. (2018), "The design, development, and implementation of student-facing learning analytics dashboards", *Journal of Computing in Higher Education*, Vol. 30 No. 3, pp. 572-598.

Buckingham Shum, S. and McKay, T.A. (2018), "Architecting for learning analytics: innovating for sustainable impact", *EDUCAUSE Review*, Vol. 53 No. 2, pp. 25-37, available at: https://er.educause.edu/articles/2018/3/architecting-for-learning-analytics-innovating-for-sustainable-impact

Butler, N. and Winne, P.H. (1995), "Feedback and self-regulated learning: a theoretical synthesis", *Review of Educational Research*, Vol. 65 No. 3, pp. 245-281.

Campbell, D.T. and Stanley, J.C. (1963), *Experimental and Quasi-Experimental Designs for Research*, Houghton Mifflin Company, Boston, MA.

Carless, D. (2007), "Learning-oriented assessment: conceptual bases and practical implications", *Innovations in Education and Teaching International*, Vol. 44 No. 1, pp. 57-66, doi: 10.1080/14703290601081332.

Carless, D. and Boud, D. (2018), "The development of student feedback literacy: enabling uptake of feedback", *Assessment and Evaluation in Higher Education*, Vol. 43 No. 8, pp. 1315-1325.

Carlson, D.C. (2015), "Try computer administered testing", in Wilson, E.J. and Hair, J.F. (Eds), *Proceedings of the 1996 Academy of Marketing Science (AMS) Annual Conference. Developments in Marketing Science: Proceedings of the Academy of Marketing Science, Springer*, pp. 264-267, doi: 10.1007/978-3-319-13144-3_84.

Conrad, D. and Openo, J. (2018), *Assessment Strategies for Online Learning: engagement and Authenticity*, Athabasca University Press, Athabasca, doi: 10.15215/aupress/9781771992329.01.

Davis, F.D. (1989), "Perceived usefulness, perceived ease of use, and user acceptance of information technology", *MIS Quarterly*, Vol. 13 No. 3, pp. 319-340.

Eggen, T.J.H.M. and Straetmans, G.J.J.M. (2000), "Computerized adaptive testing for classifying examinees into three categories", *Educational and Psychological Measurement*, Vol. 60 No. 5, pp. 713-734.

Ellis, C. (2013), "Broadening the scope and increasing usefulness of learning analytics: the case for assessment analytics", *British Journal of Educational Technology*, Vol. 44 No. 4, pp. 662-664.

Eseryel, D., Law, V., Ifenthaler, D., Ge, X. and Miller, R.B. (2014), "An investigation of the interrelationships between motivation, engagement, and complex problem solving in game-based learning", *Journal of Educational Technology and Society*, Vol. 17 No. 1, pp. 42-53, available at: www.j-ets.net/collection/published-issues/17_1

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C. and Strahan, E.J. (1999), "Evaluating the use of exploratory factor analysis in psychological research", *Psychological Methods*, Vol. 4 No. 3, pp. 272-299.

Frick, T.W. (1990), "A comparison of three decision models for adapting the length of computer-based mastery tests", *Journal of Educational Computing Research*, Vol. 6 No. 4, pp. 479-513.

Frick, T.W. (1992), "Computerized adaptive mastery tests as expert systems", *Journal of Educational Computing Research*, Vol. 8 No. 2, pp. 187-213.

Gašević, D., Greiff, S. and Shaffer, D. (2022), "Towards strengthening links between learning analytics and assessment: challenges and potentials of a promising new bond", *Computers in Human Behavior*, Vol. 134, p. 107304.

Gikandi, J.W., Morrow, D. and Davis, N.E. (2011), "Online formative assessment in higher education: a review of the literature", *Computers and Education*, Vol. 57 No. 4, pp. 2333-2351.

Heil, J. and Ifenthaler, D. (2023), "Online assessment for supporting learning and teaching in higher education: a systematic review", *Online Learning*, Vol. 27 No. 1, pp. 187-218.

Howell, J.A., Roberts, L.D., Seaman, K. and Gibson, D.C. (2018), "Are we on our way to becoming a 'helicopter university'? Academics' views on learning analytics", *Technology, Knowledge and Learning*, Vol. 23 No. 1, pp. 1-20.

Huebner, A. (2012), "Item overexposure in computerized classification tests using sequential item selection", *Practical Assessment, Research, and Evaluation*, Vol. 17, p. 12, doi: 10.7275/nr1c-yv82.

Ifenthaler, D. (2012), "Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios", *Journal of Educational Technology and Society*, Vol. 15 No. 1, pp. 38-52.

Ifenthaler, D. (2017), "Are higher education institutions prepared for learning analytics?", *TechTrends*, Vol. 61 No. 4, pp. 366-371.

Ifenthaler, D., Pirnay-Dummer, P. and Seel, N.M. (Eds) (2010), *Computer-Based Diagnostics and Systematic Analysis of Knowledge*, Springer, New York, doi: 10.1007/978-1-4419-5662-0.

Jo, I.-H., Yu, T., Lee, H. and Kim, Y. (2015), "Relations between student online learning behavior and academic achievement in higher education: a learning analytics approach", in Chen, G., Kumar, V., Kinshuk, Huang, R. and Kong, S.C. (Eds), *Emerging Issues in Smart Learning. lecture Notes in Educational Technology*, Springer, Berlin, pp. 275-287.

Johnson, W.L. and Lester, J.C. (2016), "Face-to-Face interaction with pedagogical agents, twenty years later", *International Journal of Artificial Intelligence in Education*, Vol. 26 No. 1, pp. 25-36.

Klasen, D. and Ifenthaler, D. (2019), "Implementing learning analytics into existing higher education legacy systems", in Ifenthaler, D., Yau, J.Y.K. and Mah, D.K. (Eds), *Utilizing Learning Analytics to Support Study Success*, Springer, Cham, pp. 61-72.

Könings, K.D., Seidel, T. and van Merriënboer, J.J.G. (2014), "Participatory design of learning environments: integrating perspectives of students, teachers, and designers", *Instructional Science*, Vol. 42 No. 1, pp. 1-9.

Leitner, P., Ebner, M. and Ebner, M. (2019), "Learning analytics challenges to overcome in higher education institutions", in Ifenthaler, D., Yau, J.Y.K. and Mah, D.K. (Eds), *Utilizing Learning Analytics to Support Study Success*, Springer, Cham, pp. 91-104.

Lenhart, A. (2015), *Teen, Social Media and Technology Overview 2015*, Pew Research Center, Washington, DC.

Liefeld, J.P. and Herrmann, T.F. (1990), "Learning consequences for university students using computerized mastery testing", *Educational Technology Research and Development*, Vol. 38 No. 2, pp. 19-25.

Lin, C.J. and Spray, J. (2000), *Effects of Item-Selection Criteria on Classification Testing with the Sequential Probability Ratio Test*, ACT Research Report Series, Iowa City, IA.

Matcha, W., Uzir, N.A., Gaševic̈, D. and Pardo, A. (2020), "A systematic review of empirical studies on learning analytics dashboards: a self-regulated learning perspective", *IEEE Transactions on Learning Technologies*, Vol. 13 No. 2, pp. 226-245.

Nicol, D.J. and Macfarlane-Dick, D. (2006), "Formative assessment and self-regulated learning: a model and seven principles of good feedback practice", *Studies in Higher Education*, Vol. 31 No. 2, pp. 199-218.

Pachler, N., Daly, C., Mor, Y. and Mellar, H. (2010), "Formative e-assessment: practitioner cases", *Computers and Education*, Vol. 54 No. 3, pp. 715-721.

Panadero, E., Jonsson, A. and Botella, J. (2017), "Effects of self-assessment on self-regulated learning and self-efficacy: four meta-analyses", *Educational Research Review*, Vol. 22, pp. 74-98.

Pardo, A. and Siemens, G. (2014), "Ethical and privacy principles for learning analytics", *British Journal of Educational Technology*, Vol. 45 No. 3, pp. 438-450.

Park, Y. and Jo, I.H. (2015), "Development of the learning analytics dashboard to support students' learning performance", *Journal of Universal Computer Science*, Vol. 21 No. 1, pp. 110-133, doi: 10.3217/jucs-021-01-0110.

Parshall, C.G., Spray, J.A., Kalohn, J. and Davey, T. (2002), *Practical Considerations in Computer-Based Testing*, Springer, New York.

Pokhrel, J. and Awasthi, A. (2021), "Effectiveness of dashboard and intervention design", in Sahin, M. and Ifenthaler, D. (Eds), *Visualizations and Dashboards for Learning Analytics*, Springer, pp. 93-116, doi: 10.1007/978-3-030-81222-5_5.

Quellmalz, E. (2015), "Computer-based assessment", in Gunstone, R. (Ed.), *Encyclopedia of Science Education*, Springer, Dordrecht, pp. 188-193.

Roberts, L.D., Howell, J.A. and Seaman, K. (2017), "Give me a customizable dashboard: personalized learning analytics dashboards in higher education", *Technology, Knowledge and Learning*, Vol. 22 No. 3, pp. 317-333.

Sadler, D.R. (1989), "Formative assessment and the design of instructional systems", *Instructional Science*, Vol. 18 No. 2, pp. 119-144.

Sadler, D.R. (2010), "Beyond feedback: developing student capability in complex appraisal", *Assessment and Evaluation in Higher Education*, Vol. 35 No. 5, pp. 535-550.

Sahin, M. and Ifenthaler, D. (2021), "Visualizations and dashboards for learning analytics: a systematic literature review", in Sahin, M. and Ifenthaler, D. (Eds), *Visualizations and Dashboards for Learning Analytics*, Springer, Cham, pp. 3-22, doi: 10.1007/978-3-030-81222-5_1.

Saqr, M., Fors, U. and Tedre, M. (2017), "How learning analytics can early predict under-achieving students in a blended medical education course", *Medical Teacher*, Vol. 39 No. 7, pp. 757-767.

Sarikaya, A., Correll, M., Bartram, L., Tory, M. and Fisher, D. (2018), "What do we talk about when we talk about dashboards?", *IEEE Transactions on Visualization and Computer Graphics*, Vol. 25 No. 1, pp. 682-692.

Schiefele, U., Krapp, A., Wild, K.P. and Winteler, A. (1993), "Der 'fragebogen zum studieninteresse' (FSI)", *Diagnostica*, Vol. 39 No. 4, pp. 335-351.

Schraw, G. (2010), "Measuring self-regulation in computer-based learning environments", *Educational Psychologist*, Vol. 45 No. 4, pp. 258-266.

Schuler, H. and Prochaska, M. (2001), *Leistungsmotivationsinventar*, Hogrefe, Zurich.

Schumacher, C. and Ifenthaler, D. (2018), "Features students really expect from learning analytics", *Computers in Human Behavior*, Vol. 78, pp. 397-407.

Schumacher, C. and Ifenthaler, D. (2021), "Investigating prompts for supporting students' self-regulation – a remaining challenge for learning analytics approaches?", *The Internet and Higher Education*, Vol. 49, p. 100791.

Schwendimann, B.A., Rodriguez-Triana, M.J., Vozniuk, A., Prieto, L.P., Boroujeni, M.S., Holzer, A. and Dillenbourg, P. (2016), "Perceiving learning at a glance: a systematic literature review of learning dashboard research", *IEEE Transactions on Learning Technologies*, Vol. 10 No. 1, pp. 30-41.

Shute, V.J. (2008), "Focus on formative feedback", *Review of Educational Research*, Vol. 78 No. 1, pp. 153-189.

Sitzmann, T., Ely, K., Brown, K.G. and Bauer, K.N. (2010), "Self-assessment of knowledge: a cognitive learning or affective measure?", *Academy of Management Learning and Education*, Vol. 9 No. 2, pp. 169-191.

Spray, J.A. and Reckase, M.D. (1996), "Comparison of SPRT and sequential bayes procedures for classifying examinees into two categories using a computerized test", *Journal of Educational and Behavioral Statistics*, Vol. 21 No. 4, pp. 405-414.

Tai, J.H.M., Ajjawi, R., Boud, D., Dawson, P. and Panadero, E. (2018), "Developing evaluative judgement: enabling students to make decisions about the quality of work", *Higher Education*, Vol. 76 No. 3, pp. 467-481.

Teasley, S.D., Kay, M., Elkins, S. and Hammond, J. (2021), "User-centered design for a student-facing dashboard grounded in learning theory", in Sahin, M. and Ifenthaler, D. (Eds), *Visualizations and Dashboards for Learning Analytics*, Springer, Cham, pp. 191-212, doi: 10.1007/978-3-030-81222-5_9.

Tempelaar, D.T., Rienties, B. and Giesbers, B. (2015), "In search for the most informative data for feedback generation: learning analytics in a data-rich context", *Computers in Human Behavior*, Vol. 47, pp. 157-167.

Thompson, N.A. (2007), "A practitioner's guide for variable-length computerized classification testing", *Practical Assessment, Research, and Evaluation*, Vol. 12, p. 1, doi: 10.7275/fq3r-zz60.

Tsai, N.W. (2016), "Assessment of students' learning behavior and academic misconduct in a student-pulled online learning and student-governed testing environment: a case study", *Journal of Education for Business*, Vol. 91 No. 7, pp. 387-392.

Turkay, S. and Tirthali, D. (2010), "Youth leadership development in virtual worlds: a case study", *Procedia Social and Behavioral Sciences*, Vol. 2 No. 2, pp. 3175-3179.

van der Kleij, F. and Adie, L. (2018), "Formative assessment and feedback using information technology", in Voogt, J., Knezek, G., Christensen, R. and Lai, K.W. (Eds), *International Handbook of IT in Primary and Secondary Education*, 2nd ed., Springer, Cham, pp. 601-615.

van der Linden, W.J. and Glas, C.A.W. (2000), *Computerized Adaptive Testing: Theory and Practice*, Springer, Dordrecht.

van Groen, M.M. (2012), "Computerized classification testing and its relationship to the testing goal", in Eggen, T.J.H.M. and Veldkamp, B.P. (Eds), *Psychometrics in Practice at RCEC*, RCEC, Apeldoorn, pp. 142-150.

van Groen, M.M., Eggen, T.J.H.M. and Veldkamp, B.P. (2019), "Multidimensional computerized adaptive testing for classifying examinees", in Veldkamp, B.P. and Sluijter, C. (Eds), *Theoretical and Practical Advances in Computer-Based Educational Measurement: Methodology of Educational Measurement and Assessment*, Springer, Cham, pp. 271-289, doi: 10.1007/978-3-030-18480-3_14.

Veenman, M.V.J. (2013), "Assessing metacognitive skills in computerized learning environments", in Azevedo, R. and Aleven, V. (Eds), *International Handbook of Metacognition and Learning Technologies*, Springer, New York, pp. 157-168.

Verbert, K., Govaerts, S., Duval, E., Santos, J.L., Assche, F., Parra, G. and Klerkx, J. (2014), "Learning dashboards: an overview and future research opportunities", *Personal and Ubiquitous Computing*, Vol. 18 No. 6, pp. 1499-1514.

Wald, A. (1947), *Sequential Analysis*, John Wiley, New York.

Wanner, T. and Palmer, E. (2018), "Formative self- and peer assessment for improved student learning: the crucial factors of design, teacher participation and feedback", *Assessment and Evaluation in Higher Education*, Vol. 43 No. 7, pp. 1032-1047.

Webb, M. and Ifenthaler, D. (2018), "Assessment as, for and of 21st century learning using information technology: an overview", in Voogt, J., Knezek, G., Christensen, R. and Lai, K.W. (Eds), *International Handbook of IT in Primary and Secondary Education*, 2nd ed., Springer, New York, pp. 581-600, doi: 10.1007/978-3-319-71054-9_37.

Webb, M., Gibson, D.C. and Forkosh-Baruch, A. (2013), "Challenges for information technology supporting educational assessment", *Journal of Computer Assisted Learning*, Vol. 29 No. 5, pp. 451-462.

Yoo, Y., Lee, H., Jo, I.-H. and Park, Y. (2015), "Educational dashboards for smart learning: review of case studies", in Chen, G., Kinshuk, V., Kumar, Huang, R. and Kong, S.C. (Eds), *Emerging Issues in Smart Learning*, Springer, Cham, pp. 145-155.

Corresponding author
Dirk Ifenthaler can be contacted at: dirk@ifenthaler.info