JOURNAL OF LEARNING ANALYTICS

Volume 11(3), 249–262. https://doi.org/10.18608/jla.2024.8529

Learning Analytics Dashboard Design and Evaluation to Support Student Self-Regulation of Study Behaviour

Lars de Vreugd[1], Anouschka van Leeuwen[2], Renée Jansen[3] and Marieke van der Schaaf[4]

Abstract

For university students, self-regulation of study behaviour is important. However, students are not always capable of effective self-regulation. Providing study behaviour information via a learning analytics dashboard (LAD) may support phases within self-regulated learning (SRL). However, it is unclear what information a LAD should provide, how to present information in a usable manner, and what the information's perceived usefulness is in supporting self-regulation of study behaviour. This study entails a sequential mixed design: assessing information needs in focus groups (*n*=7), exploring usability via think-aloud interviews (*n*=8), assessing usability with the System Usability Scale (*n*=42), and assessing perceived usefulness via interviews (*n*=16). Results showed that students and tutors agreed on the relevance of the constructs chosen from literature but differed in ranking the importance of new constructs. Usability exploration led to several design improvements. Perceived usefulness assessment showed the LAD supported the appraisal of study behaviour. A need for reference frames to facilitate data interpretation was vocalized. Impacts on study behaviour varied, possibly because preparatory activities were not used. Impact could be improved by further integrating the LAD into existing learning processes.

Notes for Practice

- Learning analytics dashboards (LADs) show potential in supporting student self-regulated learning (SRL), e.g., supporting the appraisal phase by presenting relevant information.
- This study details the three-phase development and evaluation process of a LAD to support student self-regulation of study behaviour and provides theoretical and practical considerations.
- Needs assessment, usability testing, and perceived usefulness evaluation are all essential elements of a LAD's final evaluation of effectiveness.
- This study's LAD supported student appraisal of study behaviour by presenting relevant information, but a lack of actionable feedback uptake led to variance in the subsequent preparatory phase.

Keywords: Learning analytics dashboard (LAD), dashboard design, self-regulated learning (SRL), higher education, needs assessment, usability test, perceived usefulness

Submitted: 19/06/2024 — **Accepted:** 29/08/2024 — **Published:** 12/10/2024

Corresponding author [1]Email: l.b.devreugd-2@umcutrecht.nl Address: Utrecht Center for Research and Development of Health Professions Education, University Medical Centre Utrecht, Heidelberglaan 100, 3508 GA Utrecht, the Netherlands. ORCID iD: https://orcid.org/0000-0003-3486-0407

[2]Email: a.vanleeuwen@uu.nl Address: Department of Education, Utrecht University, Heidelberglaan 1, 3584 CS Utrecht, the Netherlands. ORCID iD: https://orcid.org/0000-0003-2970-1380

[3]Email: r.s.jansen-14@umcutrecht.nl Address: Utrecht Center for Research and Development of Health Professions Education, University Medical Centre Utrecht, Heidelberglaan 100, 3508 GA Utrecht, the Netherlands. ORCID iD: https://orcid.org/0000-0002-8385-8322

[4]Email: m.f.vanderschaaf-4@umcutrecht.nl Address: Utrecht Center for Research and Development of Health Professions Education, University Medical Centre Utrecht, Heidelberglaan 100, 3508 GA Utrecht, the Netherlands. ORCID iD: https://orcid.org/0000-0001-6555-5320

1. Introduction

At university, students are required to self-regulate their study behaviour (Broadbent & Poon, 2015), leading to better academic performance (Schneider & Preckel, 2017) and higher levels of study satisfaction (Liborius et al., 2019). However, students are not always capable of effective self-regulation (Bjork et al., 2013). Offering students information about study behaviour

can support self-regulated learning (SRL; Marzouk et al., 2016), which can be done via a learning analytics dashboard (LAD). However, it is unclear what information needs to be included in a LAD to support student SRL. Exploring what information stakeholders deem necessary to include in a LAD likely benefits the LAD's potential to support SRL. Furthermore, how to present this required information in a usable manner, as well as a LAD's overall usability is equally important. Finally, the extent to which students perceive the presented information to be useful to support their SRL is an indicator of the LAD's potential. This paper describes the development and evaluation of a LAD in terms of needs assessment, usability, and perceived usefulness, aimed at supporting student self-regulation of study behaviour.

1.1. Self-Regulated Learning

Self-regulating students actively research what they learn and how well goals are achieved by varying approaches to learning (Winne, 2010). SRL models are of a cyclical nature and consist of at least three phases with subprocesses (Panadero, 2017). In the preparatory phase, processes such as planning (Hadwin et al., 2011), task definition (Winne, 2011), and self-motivation (Zimmerman, 2000) take place. In the subsequent performance phase, learners can strive toward achieving goals (Boekaerts & Cascallar, 2006), monitor performance (Pintrich, 2000; Hadwin et al., 2011), and apply tactics and strategies (Winne, 2011). In the appraisal phase, learners reflect on performance (Pintrich, 2000; Zimmerman, 2000) and receive or generate performance feedback (Boekaerts, 2011).

Monitoring of learning within the appraisal phase requires students to form an understanding of the current status of learning related to (personally) set goals. Discrepancies between goals and the current status may suggest actions to be undertaken (Marzouk et al., 2016). However, students are not always effective in this process, potentially caused by inabilities to manage time and study (Steel, 2007) or not choosing effective learning strategies (Bjork et al., 2013). The appraisal phase may also be hindered by learners' selective and imperfect memories about how they studied (Schunk et al., 2008). Therefore, students could benefit from support when self-regulating their learning.

1.2. Learning Analytics to Support SRL

Offering information about learning facilitates appraisal within SRL (Marzouk et al., 2016) by supporting, for example, the evaluation of performances or strategies (Roll & Winne, 2015). However, simply mirroring data to learners is often not enough to support appraisal; instead, it requires visualizing data so that learners can interpret and make sense of it (Marzouk et al., 2016). This can be done via a LAD, which offers visual displays of data to support learning (van Leeuwen et al., 2022). By making key information easily understandable, LADs increase awareness and promote reflection (Uysal & Horzum, 2021). Numerous LADs exist with varying purposes (see e.g., Valle et al., 2021). LADs differ in their manner of data visualization (Bodily & Verbert, 2017), such as visualization types that only display data, enhanced visualization types that include, for example, class averages or interactivity features, and recommender system types that provide recommendations for follow-up activities.

Within learning analytics (LA) literature, multiple LAD design methods exist. There is increasing attention to approaches for stakeholder involvement and human-centredness, utilizing relations between learning design, learning analytics, and the learning environment (Dimitriadis et al., 2021). This led to Human-Centred Learning Analytics (HCLA), where LA systems are designed with "humans" as the focal point (Buckingham Shum et al., 2024). A narrower definition of human-centred design (HCD) refers to the inclusion of users, especially during the problem definition and ideation phases of a design process (Lang & Davis, 2023). Central to participatory design is mutual learning, collective reflection-in-action, and partnership between researchers and stakeholders (Sarmiento & Wise, 2022). Co-design is a highly facilitated team-based process where researchers, developers, and teachers collaborate on designing educational innovations, realizing designs in prototypes, and evaluating prototype significance for an educational need (Wiley et al., 2024). There is consensus in the literature regarding the added value of stakeholder involvement (Lang & Davis, 2023) to incorporate relevant experiences and needs (Dollinger et al., 2019). However, relying solely on prospective user preferences is advised against, as users may not be aware of LAD possibilities or mechanisms of data collection (Matcha et al., 2019). Alfredo et al. (2024) provide numerous suggestions for HCLA, such as actively involving stakeholders (especially students), ensuring that end-user needs are met in the system's outcomes, and addressing students' passive involvement in design processes. Wiley et al. (2024) argue that expanding HCLA design approaches (among other things) positions usability and feasibility as part of the design process instead of as an endpoint.

All in all, when developing a LAD to support learners, determining the type of LAD (e.g., visualization or recommender), the phase of learning to support (e.g., goal setting or appraisal), and the target end-user (learner or teacher) should be well deliberated and substantiated. Furthermore, inclusion of stakeholder and determining which specific phases of the design process require input should be well considered. The present study describes the development of a student-facing visualization type LAD aimed at supporting self-regulation of study behaviour. An HCD approach was applied, in which we actively involved stakeholders throughout various phases of the design process to ensure that end-user (i.e., student) needs for usability and usefulness to support their self-regulation would be reflected.

1.3. LAD Design and Evaluation

In this study, we present a LAD aimed at supporting student appraisal and preparatory phases and helping them to overcome issues in self-regulation of study behaviour. The LAD's effectiveness in doing so depends on 1) whether the presented information matches student needs, 2) the perceived usability of the LAD, and 3) the perceived usefulness of the LAD.

1.3.1. Information needs

It is unclear what information a LAD requires to support student self-regulation of study behaviour. Determining information relevance is ideally based on the needs of stakeholders, including end-users (Alfredo et al., 2024; Ochoa, 2022). The main stakeholders in our study were students and tutors since they could provide relevant perspectives on what information is needed and useful for student reflection.

1.3.2. LAD usability

A LAD's success depends on its usability. After all, a lack of usability could result in end-users not using the LAD (Bodily & Verbert, 2017). Low usability negatively impacts the usefulness assessment, just as perceived ease of use influences perceived usefulness and actual system use (Technology Acceptance Model, Davis, 1989). LADs exploit human perceptual capabilities and guide decision-making by supporting or directing cognition (Yigitbasioglu & Velcu, 2012). This can be done by using graphs and colour (Park & Jo, 2019), but the underlying semantics of colour may differ per person (Al-Ayash et al., 2016). Including graphs influences usability as well, as they can lead to efficient perception (Park & Jo, 2019). Even though graphs can efficiently depict information to help with interpretation, students are prone to specific errors within each type of graph (Lai et al., 2016, p. 668). Several LAD aspects (e.g., colour and graphs) potentially influence usability and subsequent usefulness. We, therefore, explored the usability of a LAD prototype and made design improvements before re-evaluating usability and usefulness. The goal for usability in this study was to reach a level where users can understand information, can interpret scores, can navigate the LAD, and can work with the LAD fluently.

1.3.3. LAD perceived usefulness

Examining a LAD's perceived usefulness is important because these perceptions could lead to an actual effect and influence dashboard usage (Bodily & Verbert, 2017). This study's LAD aims at supporting student appraisal of study behaviour and providing preparatory information. Kirkpatrick's four-level model — Reaction, Learning, Transfer, and Result (Kirkpatrick, 1996; Kirkpatrick & Kirkpatrick, 2006) — provides a framework to evaluate usefulness sequentially, further specified by Park and Jo (2019) to the LAD context. "Reaction" can provide more specific insights into goal-orientation and information usefulness, for example. Furthermore, the framework evaluates "information usefulness," "user friendliness," and "appropriation of visual representation" (level 1, Reaction), which relate to needs assessment and LAD usability. The Kirkpatrick framework with Park and Jo (2019) sublevels will, therefore, be used for the overarching evaluation of this study's LAD.

1.4. Current Study

This study describes the development of a LAD aimed at supporting student appraisal and preparatory phases in SRL. Our aim was to outline the process of development and evaluation of the LAD, exploring information needs, usability, and usefulness to examine the LAD's potential of supporting SRL. This study consists of three design phases with corresponding research questions:

1) What information should a LAD present to support student self-regulation of study behaviour?
2) How can a LAD visualize information to support student self-regulation of study behaviour in a usable manner?
3) What is the LAD's perceived usefulness in supporting student self-regulation of study behaviour?

2. Methods

This study employs a sequential mixed design (Teddlie & Tashakkori, 2009), applying both quantitative and qualitative methods in accordance with the research questions to explore the design phases of the LAD (i.e., the Thermos dashboard; see Figure 1).

Design phase 1 was focussed on assessing stakeholder needs, resulting in a definitive list of required topics to include in the LAD. Based on focus group results, design implications were determined, and a prototype LAD was developed. Design phase 2 (academic year 2019–2020) aimed at obtaining a satisfactory level of usability. The LAD prototype's usability was therefore first explored using think-aloud interviews. Based on these results, design adjustments were made and LAD usability was then re-assessed. In design phase 3, semi-structured interviews were employed to assess the level of usefulness per Kirkpatrick level.

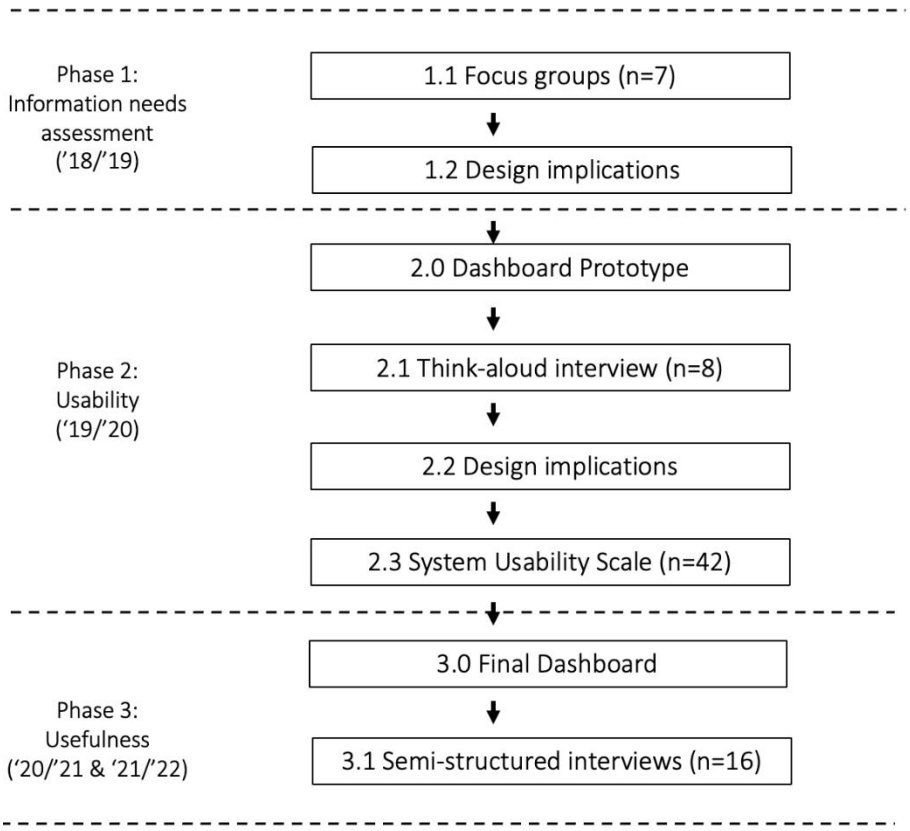

Figure 1. Overview of development phases and LAD versions.

2.1. Phase 1: Information Needs

2.1.1. Participants

Information needs were explored in two parallel focus groups in academic year 2018–2019. Students (first, second, and third year bachelor) of different study programs (e.g., Law and Medicine; total n=4) of a Dutch research university participated as future end-users. Tutors (total n=3) participated in a parallel focus group as subject matter experts. Invitations were distributed via email and flyers via multiple study programs. Participation was voluntary; all participants provided informed consent.

2.1.2. Instruments and materials

The two focus groups were organized to determine stakeholder preferences of what constructs to include in the LAD. A Nominal Group Technique (NGT) was used to enhances the value of the focus groups as information sources (Harvey & Holmes, 2012). First, participants were shown a predetermined list with potentially relevant constructs, e.g., self-efficacy (Bandura, 1997) and intrinsic goal orientation (MSLQ, Pintrich et al., 1991). They individually indicated if they expected and if they hoped these constructs would be in the LAD (both on a 7-point Likert scale). The question for students was, "*On which subjects should the learning analytics dashboard provide you with information and feedback?"* For tutors, the question was, "*On which subjects should the learning analytics dashboard provide student information, so you can offer better guidance and support your role of tutor?*" New topics could be added during the focus group.

2.1.3. Procedure

Participants first filled out the list with predetermined constructs. Next, construct importance — both original and added— was discussed. Finally, participants were asked to individually make a top five list of constructs they deemed most important to include.

2.1.4. Analysis

Means and standard deviations were calculated for the two Likert-scale questions to determine what constructs participants valued most. Ranking means from both students and tutors were calculated to determine the most important constructs overall.

2.2. Phase 2: Usability

Prototype LAD usability (Figure 2) was assessed using think-aloud interviews. After design adjustments, usability was reassessed using the System Usability Scale (SUS; Brooke, 2013).

2.2.1. Participants

For think-aloud interviews, participants (n=8, six master's students) from different programs (e.g., Educational Sciences and Language and Culture) were invited via email. All provided informed consent, and participation was voluntary. Participants received no compensation. For the SUS, participants (n=42) from various study programs voluntarily completed the survey in the LAD.

2.2.2. Instruments and materials

During think-aloud interviews, participants were shown LADs with pre-made user profiles representing various mock students (e.g., students with all "good" or all "bad" scores, different surnames, and different programs). Tasks were based on the predetermined user journey (Appendix A) and explained one at a time. For example, for the step "*interpreting presented scores,"* a task was to "*Please rank [profile name]'s top three skills/attributes, where number 1 is their strongest.*"

The SUS (Brooke, 2013) consisted of ten items on a 5-point Likert scale (1 "strongly disagree" to 5 "strongly agree"; Appendix B). Cronbach's α was .77, indicating acceptable estimated reliability (Nunnally & Bernstein, 1994). For think-aloud interviews, the LAD presented mock data in several graphs (see results Phase 1). For the SUS, a LAD was used in which design adjustments from think-aloud interviews was incorporated (see results Phase 2).

2.2.3. Procedure

For think-aloud interviews (cf. Cowan, 2019), participants individually joined an online meeting with the researcher or research assistant and repeated informed consent. Participants performed the tasks using the LAD prototype and could ask questions afterwards. All meetings were recorded (video and audio) and transcribed.

For the SUS, users were invited to complete the questionnaire through their study program and could do so via the "give feedback about the Thermos tool" option (see Figure 3).

2.2.4. Analysis

For think-aloud interviews, content analysis was applied to analyze the data. The first author and a research assistant discussed segments to assess the capability of participants to perform the user journey steps. After this deliberation, an individual "yes" or "no" judgment was made for each separate step. If the judgments did not align, the segment was discussed until agreement was achieved. For the SUS, calculated mean scores were transformed to percentages to use the Bangor et al. (2009) evaluation framework, ranging from 0–25 "worst imaginable" to 87–100 "best imaginable."

2.3. Phase 3: Perceived Usefulness

The final LAD's perceived usefulness was assessed using semi-structured interviews.

2.3.1. Participants

Students of cohort 2020–2021 (n=6) and 2021–2022 (n=10) participated from study programs that had implemented the LAD into their curriculum. All LAD integrations were aimed at students gaining awareness or insight into their study behaviours. Integrations varied from requesting LAD use before peer-to-peer reflection, as a reflection exercise, or as preparation for a one-on-one tutor conversation. Participants were required to have used the LAD to participate in the study. Participants (first, second, and third year bachelor, and first year master) from different programs (e.g., Psychology and Computer sciences) were recruited via email or during LAD explanation in study programs. Participants received €10 as compensation.

2.3.2. Instruments and materials

Interview guidelines were based on Kirkpatrick's framework, consisting of four levels (Reaction, Learning, Behaviour, Result; Appendix C). Level 4 (Result) entails changes in behaviour from level 3 (Behaviour). This level was outside of this study's scope. Research assistants trained in using the interview guidelines and procedure conducted the interviews. For usefulness assessments, the final LAD version was used (Figure 3; see results section under the design implications of phase 2).

2.3.3. Procedure

Two research assistants conducted semi-structured interviews via Microsoft Teams or face to face (whichever participants preferred). Participants could talk about their LAD results, or the research assistant could present examples of LAD results (depending on participant preference). The available LAD served as a prompt for stimulated recall. First, the researcher assistants created rapport (introduction and small talk; Cohen et al., 2018) before starting the interview. Interviews lasted 36 minutes on average (SD = 6 minutes). For online interviews, audio and video were recorded; for face-to-face interviews, only audio was recorded.

2.3.4. Analyses

Usefulness data were analyzed using Reflexive thematic analysis (Braun & Clarke, 2006; 2021). A deductive approach was applied, meaning that "*existing research and theory provide the lens through which we analyse and interpret data*" (Braun & Clarke, 2021, p. 331). The theoretical lens for data analysis was the Kirkpatrick framework in order to determine elements of the LAD's usefulness. The goal of this deductive approach was not to explore the underlying rationale but to assess the level of usefulness.

Interviews were first transcribed and pseudonymized. Data was arranged per participant, per interview question, and read by the first author and research assistant to get acquainted with the data (step 1: data familiarization). Next, initial codes were generated per participant, per interview question by the first author, and discussed with the research assistant. Codes were developed semantically, meaning that data as communicated by the participant was used to develop a code (Byrne, 2022; step 2: systematic data coding). Codes were then read and categorized in initial themes by the first author, based on their shared meaning, per Park and Jo (2019) level (e.g. "Information usefulness"). A miscellaneous category was made for codes that did not fit any category, which were discussed with the research assistant (step 3: generating initial themes). Next, the first author further developed themes by creating sub-themes that nuanced or provided explanations for the main themes. In this step, themes and codes were analyzed per Kirkpatrick level as participant answers were spread over Park and Jo (2019) levels to, for example, explain how information presentation (Park & Jo, 2019; Appropriation of visual representation) affected their perception of information usefulness (Park & Jo, 2019; Information usefulness; step 4: developing and reviewing themes).

Next, themes were further refined and defined by reviewing codes from all three Kirkpatrick levels, as participants made statements about "information interpretation." The quality and meaning of themes were reviewed in this step. For example, nine participants stated that the LAD shows useful information in general. This is relevant in answering the research question but does not provide much explanation. Subthemes were created if that theme provided more depth for a main theme (step 5: refining, defining, and naming themes). Finally, these results are reported (step 6: writing the report). During this process, several discussions and conversations were held among the authors and the research assistant to further understand and explore the data and develop themes. The analysis of interview data resulted in 480 codes and 17 main themes, with 63 subthemes.

3. Results

3.1. Phase 1: Information Needs

This phase assessed stakeholder information needs. Focus groups results of scoring predetermined constructs are presented in Table 1. Overall, students and tutors rated the importance of constructs similarly, with "fear of failure" showing the biggest difference (ΔM=1.54). Perceptions of the importance of constructs did vary within each group. For students, "Time Management" (M=6.75) was perceived as more important than "Perception of the teaching-learning environment" (M=4.13). For tutors, "Fear of Failure" (M=6.67) was deemed more important than "Self-efficacy" (M=4.33) as a potential construct for the LAD.

Table 1. Mean Scores of Tutors and Students on Construct Importance from the Predetermined List

Constructs:	Student Mean Score (1–7)	Student SD	Tutor Mean (1–7)	Tutor SD	ΔM
Concentration	4.38	0.8	5.50	0.00	1.13
Fear of failure	5.13	1.5	6.67	0.58	1.54
Test strategy	5.75	1.2	5.17	2.02	-0.58
Time management	6.75	0.3	5.67	1.04	-1.08
Engagement	4.38	1.1	5.50	1.00	1.13
Academic buoyancy	6.13	0.5	5.50	0.50	-0.63
Self-efficacy	4.88	1.1	4.33	1.26	-0.54
Perception of teaching–learning environment	4.13	1.7	4.67	1.89	0.54
Intrinsic goal motivation	5.13	1.7	5.83	0.29	0.71

Student and tutor suggestions for new constructs differed. Students suggested 1) you in relation to others (engagement and comparison), 2) mental health, 3) substance abuse, 4) perception of study progress, 5) career orientation/stimulating vision for the future, and 6) cynicism. Tutors suggested different topics: 1) decision-making skills, 2) study skills, 3) home-situation, 4) social safety net, and 5) extrinsic motivation. Tutors added 6) study strategy as an alternative to "test strategy."

Finally, Table 2 shows the ranking of the importance of each construct, with students and tutors selecting different constructs as most important. For students, two newly suggested topics made the final list of nine topics. For tutors, two new topics made the list as well, out of six topics in total. This suggests that predetermined constructs were often perceived as useful, but also shows the added value of allowing new suggestions.

Table 2. Student and Tutor Rankings of the Importance of Each Construct

	Student		**Tutor**	
	Mean	**SD**	**Mean**	**SD**
Time management*	4.25	0.96	-	-
Intrinsic goal motivation*	-	-	3.67	2.31
Mental health	-	-	3.67	0.58
Academic buoyancy*	2.50	1.91	-	-
You in relation to others	2.00	2.16	-	-
Test strategy*	1.75	2.06	-	-
Study strategy	-	-	3.67	1.53
Concentration*	1.75	2.36	-	-
Mental health/substance abuse	1.75	1.50	-	-
Decision-making skills	-	-	1.67	1.15
Fear of failure*	-	-	1.67	1.53
Engagement*	-	-	0.67	1.15
Perception of study progress	0.50	1.00	-	-
Self-efficacy*	0.25	0.50	-	-
Career orientation/stimulating vision for the future	0.25	0.50	-	-

Note: constructs that were not chosen are not shown;
an * indicates a construct from the predetermined list.

3.1.1. Phase 1 design implications

Focus group results led to a LAD prototype (Figure 2). It includes the Motivation and Engagement Wheel (MEW; Martin, 2007; Figure 2, part 3) as it relates to constructs chosen by students and tutors, such as time management, academic buoyancy, and self-efficacy. The Group Work skill graph (GSQ; Cumming et al., 2015; Figure 2, part 2) was included since it entails the construct "relation to others," chosen by students. The Study Progress widget was included since it relates to the "perception of study progress" (Figure 2, part 4). MEW and GSQ data are first gathered via a validated self-assessment questionnaire; scores per construct are averaged and then presented as percentages in the graphs. Study progress data is gathered from the university's data management system.

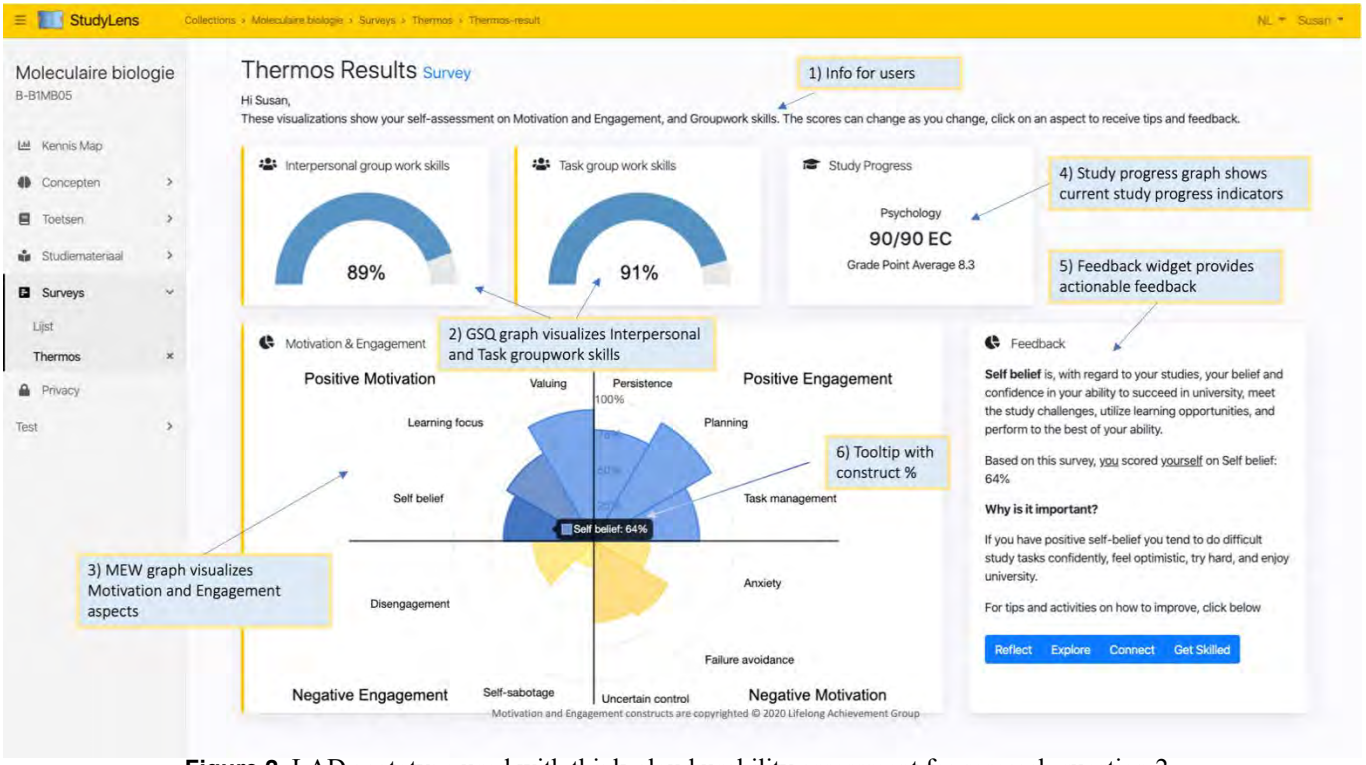

Figure 2. LAD prototype used with think-aloud usability assessment for research question 2.

The LAD aims to support users (i.e., students) to appraise aspects of their study behaviour via the MEW, GSQ, and Study Progress graphs, combined with the feedback box. The feedback box explains the meaning and importance of the constructs (Figure 2, part 5). Furthermore, the feedback box provides actionable feedback, self-improvement exercises under the "Reflect," "Explore," and "Get Skilled" buttons, and suggestions for further support (e.g., a workshop) under the "Connect" button. By presenting these options to users, the LAD aims to support the preparatory phase in which participants set goals and decide upon follow-up activities. Constructs are presented in a generic manner (i.e., not specified to a context), allowing for an overarching insight per construct, which users can apply to specific study situations. Finally, the LAD is designed to be used by students mostly of their own volition, as they are seen as agents of their own learning process. In practice, this means no teacher or tutor can monitor student results; students decide what to work on, what to share with a tutor, etc.

3.2. Phase 2: Usability

Think-aloud interviews were used to explore LAD usability, and the SUS was used to assess usability after design adjustments. Overall, results from think-aloud interviews (see Appendix D) indicated that some LAD aspects were already sufficiently usable, whereas others could be improved. For example, all participants correctly interpreted positive and negative scales (e.g., 13% is desirable for Anxiety but undesirable for Planning). All but one participant correctly compared positive to negative scales (e.g., 13% on Anxiety to 80% on Planning), although three participants had a short learning curve when comparing positive and negative scales. Determining an acceptable score was difficult; participants struggled with interpreting "*39 on a scale of 100*," for example.

LAD usability could be improved in several ways. For example, in step one (navigating empty LAD) participants mostly did not know what to do. This unclarity is undesirable and potentially demotivating, causing users to leave the LAD. Furthermore, most participants understood the goals of the exercises, but the organization of functionalities was unclear.

3.2.1. Research question 2: Design implications

Based on these findings, design adjustments were made. For example, an introduction video was added, clarifying the interpretation of positive and negative scales. Furthermore, the categorization of self-completion exercises was changed, using terms that suggest sequential steps (Prepare, Act, Reflect). Also, referrals to workshops and other supports were renamed "Additional support," and visually altered to distinguish them from self-completion exercises. This led to the LAD design, as seen in Figure 3, in which usability was assessed with the SUS.

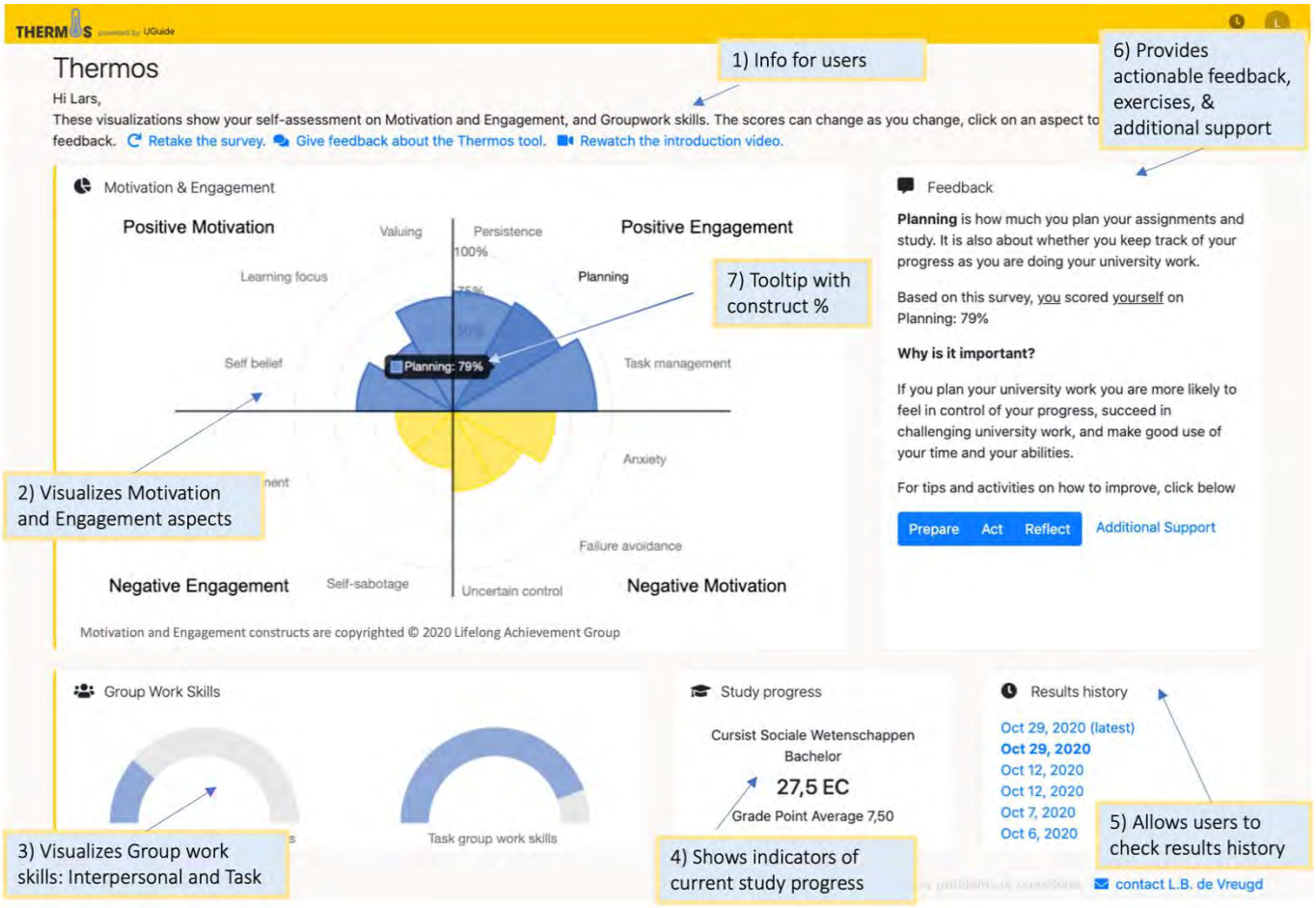

Figure 3. Final LAD version used in usability assessment (RQ 2) and usefulness assessment (RQ 3).

3.2.2. Usability reassessment

Usability reassessment with the SUS resulted in a score of M = 4.13 and SD = 0.46. This score equals 78.25%, interpretable as good–excellent usability (Bangor et al., 2009). Results showed small variance across items, indicating consistent usability. Item 1 ("I think I would like to use the Thermos dashboard frequently") scored relatively low (M = 3.00). No design adjustments were deemed necessary with this level of usability.

3.3. Phase 3: Perceived Usefulness

Findings are presented per Kirkpatrick level. The number of participants supporting a theme is presented to reflect its weight.

3.3.1. Kirkpatrick Level 1: Reaction

This level entails information usefulness and appropriation of information visualization. Participants stated that the LAD generally shows useful information (9) or potentially useful information (2), as it can provide insight into personal growth (5). Participants stated that the LAD may be especially useful for students who don't reflect much (4), e.g., participant 2 (2021–2022) stated that the LAD was useful "*for students who are not self-reflective when it comes to learning behaviour*," or participant 3 (2021–2022), who stated that it was useful "*for younger students.*" The LAD's usefulness further depends on the effort students put in (4). The potential usefulness is mostly due to the Motivation and Engagement Wheel (MEW; Figure 3, part 2); for example, participant 5 (2021–2022) mentioned that *"Having different characteristics categorized in [the MES graph] is nice; big differences provide a clear overview of what strengths and weaknesses are."* The MEW graph presents generally useful information (6), provides a good overview (2), is presented in an appropriate way (4), and is divided in a useful manner (7).

Perceived usefulness of the Group Work Skills graph (GWS; Figure 3, part 3) varied, potentially due to it being unclear (3). For example, participant 11 (2021–2022) stated that the *"GWS is a fun graph and can help a lot … (because) I find it difficult to say how I work in groups,"* but participant 8 (2021–2022) stated that "*For GWS it's not immediately clear what the value [of the information] is.*" Although the GWS graph was presented appropriately (4), the MEW presentation was

preferred (5), for example participant 10 (2021–2022) mentioned that "*It presents it just fine, but the circle graph [MES] is clearer.*" Study progress information was mostly deemed as not useful (4). For example, participant 11 (2021–2022) argued, *"I don't know what the added value of study progress could be. … I think everybody is aware of study progress and knows if they need to pass a course or not."* One participant stated that it completes the image of the student in the LAD. The actionable feedback was perceived as useful (5), potentially useful (7), or not useful (2). Some participants found the presentation appropriate (3), but most found it inappropriate (7) as there was too much text (4) or not motivating to use (3).

3.3.2. Kirkpatrick Level 2: Learning

This level entails understanding information and supporting user reflection. Interpreting the visualizations led to recognition (4) and insight into the current situation (3). For example, participant 1 said, "*I struggle with that, and that showed in the dashboard*." Visualizations also provided focus on improvement potential (2). Interpretation took some thought (7), but the provided explanatory text per construct made it easier (10). Availability of a guideline or frame of reference could facilitate interpretation (9); for example, participant 4 stated, "*but I don't know what 18% means*. *Perhaps provide some guidelines that indicate what's a good score?*" Visualization of growth or change was perceived as positive (3) as it can provide insight into progress (4). For example, participant 4 mentioned that "*At the start of this year, I scored high on anxiety; now I can see growth there; that's motivating.*" Reflection on study behaviour was impacted (10), mainly because new insights or awareness was acquired (8). For example, participant 6 said, "*It helped by providing a better picture; otherwise, I would underestimate myself.*" Some participants indicated that their reflection was not impacted (5), as they already actively reflected (5), e.g., participant 2, who said, "*I already know how to study to be successful […] could potentially have bigger effect in the first year.*" Other students forgot about the LAD (2) or put no effort into it (3). Finally, the score validity can be context dependent (2) or influenced by knowing what survey question relates to a construct (3). Also, score validity was questioned when it was 100%; e.g., participant 6 explained, "*You can never have or do something perfectly*" (2).

3.3.3. Kirkpatrick Level 3: Behaviour

This level pertains both to the motivation to change behaviour and actual changes in behaviour. Participants verbalized an impact on motivation (5); for example, participant 1 said, "*It really helped and was an extra incentive for me.*" Impact was also due to the mode of information visualization (2) and repeated use of the LAD (3). Study behaviour was impacted to some extent for some participants (5), resulting from gaining awareness (3); for example, participant 13 stated, "*Awareness has positive consequences for study behaviour.*" Most participants indicated no impact on study behaviour (11); for example, participant 3 said, "*I don't think it did [have an impact], this is just how I study.*" Lack of impact was also due to already being content with their way of studying (4), already actively reflecting on study behaviour (3), forgetting the LAD (3), or because no effort was put in (3). The impact on study behaviour could be increased with external mechanisms (3); for example, participant 5 stated, "*In my case, for example, to just once a year get an email with a link to remind you … and to see how you've progressed.*" Impact could also be increased by more accessible tips and actions (i.e., the actionable feedback; 3). A few participants indicated that the actionable feedback could impact study behaviour (2), but most participants stated that they did not use the actionable feedback (11).

4. Discussion

This study describes the development and evaluation of a LAD to support student self-regulation of study behaviour. We explored what information students and tutors deem most relevant to include in a LAD, its usability, and its perceived usefulness. We now describe the lessons learned per design phase and present several associated recommendations.

The first design phase revolved around assessing information needs. We learned that students and tutors did not differ much in scores on predetermined constructs. They did differ in their suggestions for what constructs to include, as well as which are most important, more than half of which were from the predetermined list. Even though stakeholder needs are vital in determining a LAD's content (Alfredo et al., 2024), we recommend that developers do not solely rely on stakeholder input. Providing potentially relevant information or constructs from literature allows stakeholders to weigh the potentially added value of these constructs compared to their own ideas and preferences. This results in better-informed stakeholder preferences.

The second design phase pertained to exploring and improving the LAD's usability. We learned that participants were unable to perform some steps of the user journey, e.g., navigating an empty LAD. Participants could perform most steps, such as interpreting the scales and comparing them. The usability reassessment after design adjustments was good to excellent. Interestingly, SUS item 1 ("*I think I would like to use the Thermos dashboard frequently*") scored relatively low. This could indicate that despite good usability, participants were doubtful about the LAD's usefulness. The application of think-aloud protocols to explore LAD usability provided in-depth insight on these issues. We recommend that LAD developers plan an explicit phase of usability exploration and design adjustments before evaluating a LAD's usefulness. In the development of

this LAD, our qualitative exploration provided insight into its usability and, most importantly, the reasons for its usability. This allowed us to make the necessary adjustments before evaluating its perceived usefulness.

The third design phase revolved around the perceived usefulness of the LAD in supporting the self-regulation of study behaviour. We learned that the LAD was mostly perceived to support student appraisal of study behaviour, providing new insights or awareness. The LAD did not support participants' preparatory phase, as the actionable feedback to facilitate goal setting and selection of follow-up activities were not used. Interestingly, almost all participants stated that the actionable feedback could be (potentially) useful. Finally, the perceived impact on study behaviour varied due to being satisfied with their way of studying or already actively self-reflecting. To increase a LAD's impact on learning, we recommend that developers deliberately balance the support and guidance from a LAD with integration into existing educational structures. The LAD in this study is an "enhanced visualization type" (Bodily & Verbert, 2017) and does not actively recommend activities for users to take. Uptake and use of actionable feedback were largely the student's responsibility and was influenced by perceived usefulness and the study program's integration of the LAD. Participants mentioned the necessity of putting in effort, which they may not have done (enough). Participants mentioned the potential added value of external stimuli, such as a monthly reminder or integrating LAD use into a reflection exercise. With a LAD design that actively suggests follow-up activities (i.e., a recommender type system), students are supported in taking responsibility for their learning process, which may lead to more uptake of actionable feedback. This responsibility can also (partially) lie with study advisors within a study program, advising students on what steps to take.

In this design phase, we also learned that participants sometimes struggled with establishing acceptable or passable scores, stating, for example, that *"39 out of 100 does not mean anything*." Inclusion of a reference frame to support information interpretation was suggested during both design phases two and three. We therefore recommend that LAD developers should consider including one or multiple reference frame(s) in a LAD. Reference frames facilitate data interpretation (e.g., Wise et al., 2016; de Vreugd et al., 2023) but can also elicit negative emotions (Beheshitha et al., 2016). Interestingly, participants voiced a positive perception of visualizing growth in the LAD. Long-term capturing and presenting changes in knowledge and tasks are suggested for LADs (Matcha et al., 2019), allowing users to gain context-independent insights. This may warrant the inclusion of progress reference frames in the LAD, allowing for comparison to an earlier self and evaluation of achievements or growth (Jivet et al., 2017). The results history widget (Figure 3, part 5) already serves this purpose but may require some (unwanted) additional effort, as users comparing scores requires clicking back and forth.

An overarching lesson learned was the indispensable value of including stakeholders during all design phases (see Figure 1). By doing so, we incorporated their essential perspectives as prospective end-users. This led to a LAD with perceived usefulness in supporting student appraisal of study behaviour. We, therefore, recommend LAD developers contemplate how to involve relevant stakeholders during each design phase and explicate what the added value of their involvement would be. Deliberating between different methods of stakeholder inclusion, such as participatory design or co-design, may add to this process.

5. Future Research

Future research could investigate the impact of LAD use on Kirkpatrick level 4 (the results from behavioural changes in level 3) and how LAD design contributes to this level. If a LAD supports the appraisal of study behaviour, this may lead to changes in study behaviour, which in turn may result in better academic performance or reduced drop-out. Future research should examine this final level via longitudinal studies in authentic educational contexts.

Future research should also investigate the gap between a LAD's support of appraisal and lack of impact on study behaviour, and how LAD design may mitigate this gap. This may be due to students' lack of goal setting and intention to achieve goals, which is part of the preparatory phase in SRL. This research may reveal explanatory factors between the LAD's support of appraisal and lack of impact on study behaviour.

6. Conclusion

The LAD in this study supported student appraisal of study behaviour. The preparatory phase was not supported since LAD elements were not used, despite the perceived potential usefulness of those elements. The impact on study behaviour from LAD use varied, for which several reasons were given by participants.

Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors declared no financial support for the research, authorship, and/or publication of this article.

Acknowledgments

V. Numan – data collection (research assistant); Q. Maxwell – data collection (research assistant); L. van Kesteren – data analysis support (research assistant).

References

Al-Ayash, A., Kane, R. T., Smith, D., & Green-Armytage, P. (2016). The influence of color on student emotion, heart rate, and performance in learning environments. *Color Research & Application*, *41*(2), 196–205. https://doi.org/10.1002/col.21949

Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Human-centred learning analytics and AI in education: A systematic literature review. *Computers and Education: Artificial Intelligence*, *6*, 100215. https://doi.org/10.1016/j.caeai.2024.100215

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.

Bangor, A., Miller, J., & Kortum, P. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, *4*(3), 114–123. https://dl.acm.org/doi/10.5555/2835587.2835589

Beheshitha, S. S., Hatala, M., Gašević, D., & Joksimović, S. (2016). The role of achievement goal orientations when studying effect of learning analytics visualizations. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 54–63). ACM Press. https://doi.org/10.1145/2883851.2883904

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417–444. https://doi.org/10.1146/annurev-psych-113011-143823

Bodily, R., & Verbert, K. (2017). Review of research on student-facing learning analytics dashboards and educational recommender systems. *IEEE Transactions on Learning Technologies*, *10*(4), 405–418. https://doi.org/10.1109/TLT.2017.2740172

Boekaerts, M., & Cascallar, E. (2006). How far have we moved toward the integration of theory and practice in self-regulation? *Educational Psychology Review*, *18*, 199–210. https://doi.org/10.1007/s10648-006-9013-4

Boekaerts, M. (2011). Emotions, emotion regulation, and self-regulation of learning. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 408–425). Routledge.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, *18*(3), 328–352. https://doi.org/10.1080/14780887.2020.1769238

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education*, *27*, 1–13. https://doi.org/10.1016/j.iheduc.2015.04.007

Brooke, J. (2013). SUS: A retrospective. *Journal of Usability Studies*, *8*(2), 29–40.

Buckingham Shum, S., Martínez-Maldonado, R., Dimitriadis, Y., & Santos, P. (2024). Human-centred learning analytics: 2019–24. *British Journal of Educational Technology*, *55*(3), 755–768. https://doi.org/10.1111/bjet.13442

Byrne, D. (2022). A worked example of Braun and Clarke's approach to reflexive thematic analysis. *Quality & Quantity*, *56*(3), 1391–1412. https://doi.org/10.1007/s11135-021-01182-y

Cohen, L., Manion, L., & Morrison, K. (2018). Interviews. In L. Cohen, L. Manion, & K. Morrison (Eds.), *Research methods in education* (8th ed., pp. 506–541). Routledge. https://doi.org/10.4324/9781315456539

Cowan, J. (2019). The potential of cognitive think-aloud protocols for educational action-research. *Active Learning in Higher Education*, *20*(3), 219–232. https://doi.org/10.1177/1469787417735614

Cumming, J., Woodcock, C., Cooley, S. J., Holland, M. J. G., & Burns, V. E. (2015). Development and validation of the groupwork skills questionnaire (GSQ) for higher education. *Assessment & Evaluation in Higher Education*, *40*(7), 988–1001. https://doi.org/10.1080/02602938.2014.957642

de Vreugd, L., Jansen, R., van Leeuwen, A., & van der Schaaf, M. (2023). The role of reference frames in learners' internal feedback generation with a learning analytics dashboard. *Studies in Educational Evaluation*, *79*, 101303. https://doi.org/10.1016/j.stueduc.2023.101303

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319–340. https://doi.org/10.2307/249008

Dimitriadis, Y., Martínez-Maldonado, R., & Wiley, K. (2021). Human-centered design principles for actionable learning analytics. In T. Tsiatsos, S. Demetriadis, A. Mikropoulos, & V. Dagdilelis (Eds.), *Research on e-learning and ICT in education: Technological, pedagogical and instructional perspectives* (pp. 277–296). Springer Cham. https://doi.org/10.1007/978-3-030-64363-8_15

Dollinger, M., Liu, D., Arthars, N., & Lodge, J. (2019). Working together in learning analytics towards the co-creation of value. *Journal of Learning Analytics*, *6*(2), 10–26. https://doi.org/10.18608/jla.2019.62.2

Hadwin, A. F., & Järvelä, S. (2011). Self-regulated, co-regulated, and socially shared regulation of learning. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 64–84). Routledge.

Harvey, N., & Holmes, C. A. (2012). Nominal group technique: An effective method for obtaining group consensus. *International Journal of Nursing Practice*, *18*(2), 188–194. https://doi.org/10.1111/j.1440-172X.2012.02017.x

Jivet, I., Scheffel, M., Drachsler, H., & Specht, M. (2017). Awareness is not enough: Pitfalls of learning analytics dashboards in the educational practice. *Proceedings of the 12th European Conference on Technology Enhanced Learning* (EC-TEL 2017), 12–15 September 2017, Tallinn, Estonia (pp. 82–96). Springer Cham. https://doi.org/10.1007/978-3-319-66610-5_7

Kirkpatrick, D. (1996). Great ideas revisited. *Training & Development*, *50*(1), 54–60.

Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating training programs: The four levels* (3rd ed.). Berrett-Koehler Publishers.

Lai, K., Cabrera, J., Vitale, J. M., Madhok, J., Tinker, R., & Linn, M. C. (2016). Measuring graph comprehension, critique, and construction in science. *Journal of Science Education and Technology*, *25*, 665–681. https://doi.org/10.1007/s10956-016-9621-9

Lang, C., & Davis, L. (2023). Learning analytics and stakeholder inclusion: What do we mean when we say "human-centered"? *Proceedings of the 13th International Conference on Learning Analytics and Knowledge* (LAK '23), 13–17 March 2023, Arlington, TX, USA (pp. 411–417). ACM Press. https://doi.org/10.1145/3576050.3576110

Liborius, P., Bellhäuser, H., & Schmitz, B. (2019). What makes a good study day? An intraindividual study on university students' time investment by means of time-series analyses. *Learning and Instruction*, *60*, 310–321. https://doi.org/10.1016/j.learninstruc.2017.10.006

Martin, A. J. (2007). Examining a multidimensional model of student motivation and engagement using a construct validation approach. *British Journal of Educational Psychology*, *77*(2), 413–440. https://doi.org/10.1348/000709906X118036

Marzouk, Z., Rakovic, M., Liaqat, A., Vytasek, J., Samadi, D., Stewart-Alonso, J., Ram, I., Woloshen, S., Winne, P. H., & Nesbit, J. C. (2016). What if learning analytics were based on learning science? *Australasian Journal of Educational Technology*, *32*(6). https://doi.org/10.14742/ajet.3058

Matcha, W., Uzir, N. A., Gašević, D., & Pardo, A. (2019). A systematic review of empirical studies on learning analytics dashboards: A self-regulated learning perspective. *IEEE Transactions on Learning Technologies, 13*(2), 226–245. https://doi.org/10.1109/TLT.2019.2916802

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Ochoa, X. (2022). Multimodal learning analytics: Rationale, process, examples, and direction. In C. Lang, A. F. Wise, A. Merceron, D. Gašević, & G. Siemens (Eds.), *The handbook of learning analytics* (2nd ed., pp. 54–65). https://doi.org/10.18608/hla22.006

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, *8,* 422. https://doi.org/10.3389/fpsyg.2017.00422

Park, Y., & Jo, I.-H. (2019). Factors that affect the success of learning analytics dashboards. *Educational Technology Research and Development*, *67*(6), 1547–1571. https://doi.org/10.1007/s11423-019-09693-0

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). Academic Press. https://doi.org/10.1016/B978-012109890-2/50043-3

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ).* National Center for Research to Improve Postsecondary Teaching and Learning.

Roll, I., & Winne, P. H. (2015). Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics*, *2*(1), 7–12. https://doi.org/10.18608/jla.2015.21.2

Sarmiento, J. P., & Wise, A. F. (2022). Participatory and co-design of learning analytics: An initial review of the literature. *Proceedings of the 12th International Conference on Learning Analytics and Knowledge* (LAK '22), 21–25 March 2022, Online (pp. 535–541). ACM Press. https://doi.org/10.1145/3506860.3506910

Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin*, *143*(6), 565–600. https://doi.org/10.1037/bul0000098

Schunk, D. H., Pintrich, P. R., & Meece, J. L. (2008). *Motivation in education: Theory, research, and applications* (3rd ed.). Pearson.

Steel, P. (2007). The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological Bulletin*, *133*(1), 65–94. https://doi.org/10.1037/0033-2909.133.1.65

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. SAGE Publications.

Uysal, M., & Horzum, M. B. (2021). Designing and developing a learning analytics dashboard to support self-regulated learning. In M. Sahin & D. Ifenthaler (Eds.), *Visualizations and dashboards for learning analytics* (pp. 477–496). Springer Cham. https://doi.org/10.1007/978-3-030-81222-5_22

Valle, N., Antonenko, P., Dawson, K., & Huggins-Manley, A. C. (2021). Staying on target: A systematic literature review on learner-facing learning analytics dashboards. *British Journal of Educational Technology*, *52*(4), 1724–1748. https://doi.org/10.1111/bjet.13089

van Leeuwen, A., Teasley, S. D., & Wise, A. F. (2022). Teacher and student facing learning analytics. In C. Lang, A. F. Wise, A. Merceron, D. Gašević, & G. Siemens (Eds.), *The handbook of learning analytics* (2nd ed., pp. 130–140). https://doi.org/10.18608/hla22.013

Winne, P. H. (2010). Bootstrapping learner's self-regulated learning. *Psychological Test and Assessment Modeling*, *52*(4), 472–490.

Winne, P. H. (2011). A cognitive and metacognitive analysis of self-regulated learning. In D. H. Schunk & B. Zimmerman (Eds.), *Handbook of self-regulation of learning and performance* (pp. 29–46). Routledge.

Wiley, K., Dimitriadis, Y., & Linn, M. (2024). A human-centred learning analytics approach for developing contextually scalable K–12 teacher dashboards. *British Journal of Educational Technology*, *55*(3), 845–885. https://doi.org/10.1111/bjet.13383

Wise, A. F., Vytasek, J. M., Hausknecht, S., & Zhao, Y. (2016). Developing learning analytics design knowledge in the "middle space": The student tuning model and align design framework for learning analytics use. *Online Learning*, *20*(2), 155–182.

Yigitbasioglu, O. M., & Velcu, O. (2012). A review of dashboards in performance management: Implications for design and research. *International Journal of Accounting Information Systems*, *13*(1), 41–59. https://doi.org/10.1016/j.accinf.2011.08.002

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic Press. https://doi.org/10.1016/b978-012109890-2/50031-7

Appendix A: User Journey Steps for Successful LAD Use

Thermos dashboard User Journey	Foci:	Successful user behaviour:
0. Log in with University ID		N/A
User is presented with a blank dashboard, which includes instructions on how to take the survey and how to see their scores.	1. See empty data visualization	Participants know: • that taking the survey will provide them insight into their motivation, engagement, group work skills, and study progress. • when they complete the survey that they will obtain a data visualization which can be inspected for insight how to find the "survey" button.
User fills out the questionnaire.	2. Fill out survey	Participants understand how to complete the survey.
User receives initial feedback in the form of a data visualization with text, thereby gaining insight into their study-related behaviour.	3. See data visualized, interpret scores, gain insight into behaviour, and decide what to work on	Participants: • can determine how well a student is doing in their motivation, engagement, group work skills, and study progress. • can recognize that there are negatively and positively weighted constructs e.g., by pointing out the blue and yellow colours, or "positive" and "negative" qualifiers. • determine if a student needs to improve on anything to do better in their study program e.g., they inspect their study progress data and compares to the construct scores. • correctly determine which constructs a student should improve on e.g., they understand how to interact with the positive and negative weights of the groups. • correctly determine which constructs a student performs best on e.g., they understand how to interact with the positive and negative weights of the groups.
User decides, based on the text and data visualization, if and what constructs needs to be worked on to do better in their study program, are interesting, or important.		
User decides how to work on the constructs identified in step 4.	4. Using visual or textual feedback	Participants can: • explain how a student can improve on a particular construct e.g., by referencing the links or text suggestions provided in the dashboard. • establish a learning strategy for future development e.g., listing several suggestions to be followed or an order of activities.
User interacts with interface provided exercises that develop said constructs.	5. Using the follow-up exercises	Participants: • access the provided follow-up exercises. • understand how the follow-up exercises are organized. • integrate the provided activities in a learning plan/strategy for improvement.

Appendix B: System Usability Scale

Item	Statement	M	SD
1	I think that I would like to use the Thermos dashboard frequently.	3.00	1.05
2	I found the Thermos dashboard unnecessarily complex (-).	1.65	0.59
3	I thought the Thermos dashboard was easy to use.	4.32	0.58
4	I think that I would need the support of a technical person to use the Thermos dashboard.	1.35	0.48
5	I found the various functions in the Thermos dashboard were well integrated.	4.05	0.57
6	I thought there was too much inconsistency in the Thermos dashboard.	2.03	0.60
7	I would imagine that most people would learn to use the Thermos dashboard very quickly.	4.35	0.54
8	I found the Thermos dashboard very cumbersome (awkward) to use.	1.65	0.68
9	I felt very confident using the Thermos dashboard.	3.86	0.67
10	I needed to learn a lot of things before I could get going with the Thermos dashboard.	1.59	0.69

Appendix C: Interview Guideline

KP1: Reaction	P&J: Information usefulness
	"To what extent did the dashboard show information you wanted to know?"[1]
	"To what extent do you feel the visualizations present useful information for you?"[2]
	"To what extent do you feel the visualizations present useful information for students in general?"[2]
	"To what extent do you feel the content of actionable feedback can help improve study behaviour?"[2]
	P&J: Appropriation of visual representation
	"To what extent do you feel the visualizations are presented in a usable manner for you?"[2]
	"To what extent do you feel the visualizations are presented in a usable manner for students in general?"[2]
	"To what extent do you feel the presentation of the actionable feedback can help you improve study behaviour?"[2]
KP2: Learning	P&J: Understanding
	"How did you interpret the visualizations, and can you give an example?"[1]
	"To what extent did you think it was easy to understand the visualizations?"[1]
	P&J: Reflection
	"To what extent did the dashboard help to reflect on study behaviour?"[1]
	"To what extent did dashboard use change how conscious you are about your study behaviour?"[2]
KP3: Transfer	P&J: Learning motivation
	"Did the dashboard motivate you to work on study behaviour? And why?"[1]
	P&J: Behavioural change
	"Did your study behaviour change from using the dashboard?"[1]
	"Did you take subsequent steps after using the dashboard?"[1]
	"To what extent did dashboard use change the amount or way of reflecting on your study behaviour?"[2]
	"To what extent do you feel using the dashboard changed something in the way you study?"[2]
	"To what extent did you use the actionable feedback, and did it change how you study?"[2]

Note: Superscript numbers [1] and [2] indicate that the question was asked in the 2020–2021 or 2021–2022 evaluation, respectively.

Appendix D: Overview of Usability Steps and Illustrating Examples

Analysis questions:	No/Yes	Example
User step: Navigate empty dashboard		
1.1 Does this give the user a sense of what the tool is about before filling out the questionnaire?	7 / 1	P3 (No): "What is the next, the, the function of this page… It's not very obvious."
1.2 Is it clear that users should take the survey?	4 / 4	P2 (Yes): "I think something needs to be filled out." P7 (No): "The matrix needs information, but I don't know how."
1.3 Is it clear how users can take the survey?	6 / 2	P1 (No): "The current survey link is too inconspicuous, needs to be larger."
User step: Fill out the survey		
2.1 Do users read the explanatory text for the survey?	1 / 7	P7 (Yes): "This is an explanation for the questions."
User step: Interpret data visualizations		
3.1 Are users able to interpret the positive scales correctly?	0 / 8	P6 (Yes) made all correct interpretations and was able to compare constructs.
3.2 Do users favour positive constructs for the first task? (Identify profile's biggest strength)	4 / 4	P4 (No) compared a positive construct score to a negative construct score. P5 (Yes) Only considered positive constructs to as strengths.
3.3 Are users able to interpret the negative scales correctly?	0 / 8	P1 (Yes) "Chunks in yellow on the bottom represent what needs improvement."
3.4 Are users able to compare positive and negative scales like we want them to?	1 / 7	P3 (Yes) "having a high score on self-belief is better than on disengagement."
3.5 Were all users able to compare positive and negative scales during the first task?	3 / 5	P3 (No) Consistently compared percentages, regardless of the construct being positive or negative. P1 (No) interpreted incorrectly during first task, realized the mistake during second task.
3.6 Do users refer to all components of the tool?	0 / 8	P8 (Yes) used all graphs for interpretations and to get an overall idea.
3.7 Do users have a sense of how well the fictitious students are doing?	1 / 7	P5 (No) does not relate constructs, only focuses on the MES graph. P2 (Yes) sees Hakim's not feeling well at school (high neg. MES scores), but no one would see because he has his EC's and does well in groups.
3.8 Do users establish a "passable" or acceptable score?	8 / 0	P2 (No) "I don't know if this is good, 39 on a scale of 100, is that a lot? Is that normal in relation to others? When I get this, am I doing good or not?"
User step: Decide what construct to explore and what to work on		
4.1 Do users read the text in the feedback box?	0 / 8	P6 (Yes) Several times, for clarification and to rank constructs.
4.2 Do users click the links for exercises or additional support?	2 / 6	P3 (No) read the description for Planning, then stated, "So they don't give any tips here." P4 (Yes) said he would use them to advise the student.
4.3 Do users use information from the dashboard e.g., construct text or provided exercises to advise the fictitious student?	2 / 6	P5 (No) based her advice on her personal opinions. P1 (Yes) used construct explanations, tips, and activities on how to improve.
User step: Decide how to work on that construct		
5.1 Do users understand how the activities are organized?	8 / 0	P1 (No) "I don't know which one would help me most of these four [buttons]."
5.2 Do users click on the follow-up exercise links?	2 / 6	P8 (Yes) to explore what is available.
5.3 Do users reference these activities when completing tasks?	4 / 4	P5 (No) Does not reference these tasks. P1 (Yes) knows these are meant as support but doesn't understand which one would be best.