How do visualizations and automated personalized feedback engage professional learners in a Learning Analytics Dashboard?

Sarah Alcock
Institute of Educational Technology, The Open University
sarah.alcock@open.ac.uk

Maria Aristeidou
Institute of Educational Technology, The Open University
maria.aristeidou@open.ac.uk

Bart Rienties
Institute of Educational Technology, The Open University
bart.rienties@open.ac.uk

Soraya Kouadri Mostéfaoui
Computing & Communication, The Open University
soraya.kouadri@open.ac.uk

ABSTRACT
Learning Analytics Dashboards (LAD) are the subject of research in a multitude of schools and higher education institutions, but a lack of research into learner-facing dashboards in professional learning has been identified. This study took place in an authentic professional learning context and aims to contribute insights into LAD design by using an academic approach in a practice-based environment. An existing storytelling LAD created to support 81 accountants was evaluated using Technology Acceptance Model, finding a learner expectation for clarity, conciseness, understanding and guidance on next steps. High usage levels and a ‘take what you need’ approach was identified, with all visualizations and automated personalized feedback being considered useful although to varying degrees. Professional learners in this study focus on understanding and acting upon weaknesses rather than celebrating strengths. The lessons for LAD design are to offer choice and create elements which support learners to take action to improve performance at a multitude of time points and levels of success.

CCS CONCEPTS
• Applied computing → Education; Computer-assisted instruction; • Human-centered computing → Visualization.

KEYWORDS
Learning Analytics Dashboard, LAD, professional learning, Technology Acceptance Model, assessment, feedback, data storytelling, personalization, accountancy

1 BACKGROUND
1.1 Context and aims of study
This paper focuses on Learning Analytics Dashboards (LAD), a distinct segment of Learning Analytics (LA) seeking to provide learning insights by visualizing the trace data produced in an online environment with the express objective of improving learning [23]. It applies academic research processes and rigor to a practitioner environment to examine perceptions of, engagement with, and use of a learner-facing LAD in a professional learning context rather than the higher education context in which most LAD research is situated [33, 34]. The paper therefore addresses existing research gaps including professional learning being overlooked by researchers [3], the need for better links between research and practice [13], calls for the LA research community to remember its origins and utilize learner data in context for research [26] and the overall need for more research into learner-facing dashboards [10].

1.2 Challenges in LAD design
LAD are used widely in classrooms and online teaching programs to present all manner of data to all manner of audiences. This can include the immediate display of classroom activity scores to a teacher [9, 24], learner-facing dashboards supporting engagement and learning [22, 37] and dashboards making predictions of success based on past behavior [8, 29]. As LAD have evolved from the initial excitement of access to and display of data trails [25], merely presenting behavior data (e.g., frequency of log-ins or minutes of videos watched) has become simplistic and banal. To be truly useful, learner-facing LAD must present appropriate data to support learning within the given context [17], and do so in a transparent, trustworthy and effective way. Useful LAD do not simply describe the past; they ‘close the loop’ [7], direct towards action [39] and tell the story of the learning data [12, 14].

Despite the abundant applications and stakeholders – or perhaps because of them – identifying the best practice ‘how’ to design an effective LAD has not yet been achieved [17]. Attempts have been made to set out standards for LADs, for example a conceptual framework for choosing visualizations based on, among other things, educational and visualization concepts [35]. However, the resulting elaborate table of suggestions and considerations is challenging to follow without thorough understanding of the underlying theories and is certainly not a starting point for a beginner. Authors suggest next steps might include development of a system to automatically recommend visualization choice based on user input; LADStudio...
[31] does just that. However, these fixed and rule-based systems cannot possibly take into account every context and every situational requirement as acknowledged by [35] in their conclusions – “Often different combinations might be needed depending on different learning contexts and needs” (p.37). There is a continuing need then to explore LAD design principles with a focus on supporting learning.

1.3 Beyond the numbers in a LAD

Visualizations of data are a cornerstone of LAD, but the problem of data literacy – the ability to interpret numbers and graphs remains pervasive. [5] emphasized elements of personalization and choice within their LAD, giving learners the agency to select inputs and query datasets to create a LAD data visualization of their preference. Using the Technology Acceptance Model [8], authors found 28 of 38 (74%) participants (bachelor and master level computer science students) agreed they would use the tool frequently during future courses (Perceived Usefulness, PU), however it is notable that only 13 of 38 (34%) agreed the LAD was easy to use (Perceived Ease of Use, PEU). This mismatch highlights a burden on the user; will they really invest time to learn the system and make use of it during their ongoing studies? Are they data literate enough to make sensible choices and useful interpretations? Solutions to this issue include the need for specific training [27], such as those implemented for tutors by [20]. However, this project was not fully successful in creating impactful change for all the tutors involved, so success in similar training for learners seems ambitious at best. Learners, particularly professional learners who are also employed with significant responsibilities, do not have time.

Instead, we return to the notion of educational data storytelling [12]. Telling the story of the LAD data requires words – explanations, scaffolding and advice – as well as data visualizations. Feedback on learner outputs is known to be “one of the most powerful influences on learning and achievement” [19, p.81], and the problem of providing automated personalized feedback to a large cohort - previously thought to be impossible [4, 22] - has become a mainstream research topic [6]. Furthermore, recent research argues it is insufficient for a LAD tool to be created independently from the course or learning context. For coherence and seamless integration, LAD need to be designed with key stakeholders using techniques such as embedded design teams [1] or co-creation [11]. Key stakeholders include managers or administrators with an overall view of institutional learning aims. However, they should be led by the needs of tutors and, particularly for the type of LAD in this research, the needs of learners. Integration with learning design, therefore, becomes crucial; the LAD tool must be embedded within the course, not created as an addendum that, at best, is ignored or, at worst, becomes another time-consuming job for the educator [36] with no demonstrable added value to learning. The tutor-designed, learner-facing LAD in this research seems to answer these calls.

1.4 Research aims

The paper aims to make a unique contribution to LAD research by evaluating the Perceived Ease of Use (PEU), Perceived Usefulness (PU) and reported use of a learner-facing assessment LAD tool in professional education using TAM. The study was carried out in a commercial accountancy learning and assessment environment with professional learners operating at a postgraduate level, carrying the burden of work-based responsibilities alongside preparing for a high-stakes examination for admission to the Institute of Chartered Accountants in England and Wales (ICAEW). If passed, they achieve the Associate Chartered Accountant (ACA) designation and all the career opportunities that come with it, but failing the exam puts their employment at risk. There are 580,000 professional accountancy learners worldwide studying for membership in UK-based chartered accountancy bodies, with 160,000 learners based in the UK [16] where this investigation took place. With such substantive numbers and the potential to expand research and results across other business and professional qualifications, the importance of this study is irrefutable.

A learner-facing assessment LAD was embedded within a four-month exam preparation program and took what could be described as a “storytelling” [12] approach to the data presented. This fine-grained mixed methods study examined professional learner engagement with the LAD for 81 professional accountants working towards their final qualifying exam over four months. Tables, graphical visualizations, written explanations and automated personalized feedback target impact on learning through a variety of data display and storytelling approaches.

This study aims to investigate professional learners’ expectations of and use of a LAD as they prepare for a high-stakes exam through the following Research Questions (RQ):

RQ1: How useful do professional learners expect the Learning Analytics Dashboard (LAD) to be for revision?

RQ2: How did the professional learners use the LAD, and, if applicable, why was it helpful to their revision?

2 METHODOLOGY

2.1 The professional learning context

This study intended to explore professional learners’ perceptions and use of LAD. It was conducted in an authentic learning context, answering calls for practice-based evidence for the effects of a LA tool on learning [12, 40]. Professional accountancy tuition companies operate in a competitive market where high pass rates and a positive learning experience are vital for a reputation of excellence and (therefore) continuing commercial success. The LAD examined in this paper was designed to answer a specific need in preparing learners for the ICAEW Case Study exam. Crucially, the exam does not test knowledge but rather skills-based application of knowledge to a business scenario; the exam is designed to reflect the professional world learners are operating in, bridging the gap between formal and work-based learning. Overall, there are 200 marks available, and learners must achieve 100 or more (50%) to pass. It is not considered possible to achieve all the available marks in the exam time, but they are presented to account for differing ways to answer the requirements. An individual marking key is created for each exam. The need for tutors to explain mock exam results has driven development of the mock exam result LAD. The LAD presents visualizations of marking data and appropriate feedback written by the tutor, selected for inclusion based on the input of marks awarded (or, more crucially, the marks not awarded).
Professional accountancy learners in this program used at least six mock exams over the four-month tuition period. Three mock exams were professionally marked, with two of these using the same exam scenario as the real ICAEW exam. The tuition program was a blended course, with seven days of classroom teaching and an expectation of at least the same time again spent working independently. Each marking key was created in MS Excel, and once exam marking was completed, it automatically generated the mock exam results and LAD output. These were sent to the professional learner as an email attachment outside of class time, with learners expected to review their performance independently, making improvement plans. This simple technology was chosen for agility and budgetary reasons; ICAEW review exams and syllabi annually, and tutors must adjust the LAD design accordingly. Development of a specific online tool would be costly, as would maintenance, but all professionals – tutors, markers and learners - are proficient in using Excel in their daily practice.

2.2 The LAD

The LAD in this context was a non-academic tutor (practitioner) led project, rather than a developer-led project, meaning the design was informed by an experienced tutor’s understanding of the very specific needs of ICAEW Case Study learners. No specific learning design principles, other than to explain the data and next steps, were overtly used however, as described below, educational data storytelling attributes are evident. The Case Study marking key was built in MS Excel, and the LAD was created through the extensive and clever use of formulae to automatically generate personalized data visualizations and written feedback from the standard marking process. This study focused on six elements of the LAD; three visualizations, one standardized text guidance and two personalized text elements, as shown in Figures 1 through 6.

Element 1 provided a graphical overview of the exam performance, showing how the four requirement scores (ES – Executive Summary, R1, R2, R3 – Requirement 1 through 3) made up the total result and whether the pass threshold was met. It was the introduction to the story, setting up context. Element 2 contained the detailed numbers, by requirement (including OAR – Overall Assessment Requirements) and by skills (AUI – Assimilating and Using Information, SPS – Structuring Problems and Solutions, AJ – Applying Judgement, CR – Conclusions and Recommendations), as well as totals – everything learners needed to know, but it might be considered complex. Colors were added for clarity, to draw the learner’s eye to the strengths and weaknesses of their work. The detail could be overwhelming, particularly for a low-scoring result, so the fixed text in element 3 took an overview approach to ‘what now’ – reminding learners of the fundamental skills and tasks that could address issues. Next, learners delved into their numerical results by requirement and then by skills in the form of a tabular visualization of the marking key (element 4). Again, the green-to-red coloring helped learners to identify where to look further for problems to address. Before they did this, element 5 offered a written performance summary, reminding them how to proceed. At the most detailed level, element 6 gave specific advice about singular marking points, including praise when something is done well. Thus, the LAD told the individual story of every learner, through every mock exam. It is worth noting that green is a dominant color throughout Figures 2 – 6, even in ‘neutral’ areas of the LAD – this is the tuition company’s brand color, so the professional learners were familiar with tuition material containing tables and headings using this color.
2.3 Data collection

The study spanned the four-month tuition program which included classroom and home study. Participants were recruited from four tuition centers, and class gender and age demographics were typical for the Case Study exam, aligned with published statistics [16], p10, p12. More importantly, participants were all accountancy learners with a minimum of 24 months of practical work experience preparing for the last examination of a qualification that would give them the ‘chartered’ designation so critical to progress in their chosen profession. 81 professional learners answered Self Reporting Questionnaire 1 (SRQ1, related to RQ1) in week 1 of their study program, and seventy-nine completed Self Reporting Questionnaire 2 (SRQ2, related to RQ2) in week 14, a response rate of 65%. Learners were based in London and/or the southwest of the UK. Appropriate ethics approval was gained and there were no withdrawals. A key challenge of data collection was to do so promptly and authentically without interfering with or distracting participants from their exam preparation tasks. The research took a mixed methods approach, using classroom observations and self-reporting questionnaires (with Likert scales and free text options). Principles of the Technology Acceptance Model (TAM2), [38], and Perceived Usefulness (PU) / Perceived Ease of Use (PEU) were applied to the six specific elements of the LAD. A shorter version of TAM was applied to limit questionnaire fatigue as participants had to rate each of the six LAD elements with four questions each (i.e., How easy is it to understand? (Likert scale 1-5) Why? (free text) and How likely are you to use it to understand your performance (Likert scale 1-5), Why? (free text). As each LAD was a separate MS Excel file, no data logs of when and for how long the LAD was accessed were available. Therefore, in-class self-reporting measures were the least intrusive method for gathering data from already busy and stressed learners. Time was allocated within the class to answer paper-based questionnaires, and many candidates took the opportunity to answer the free text ‘why’ element for at least some of the items. It was considered important to glean a range of both quantitative and qualitative data to gain a depth of understanding [15], and to gain insight into perceptions and use of the LAD [21].

In line with recommendations [35], the first questionnaire was administered in class during week 1 and focused on how (a) easy to understand (PEU) and (b) how likely participants were to use (PU) the six elements of the LAD in revision (RQ1). It included 12 closed items (e.g., How likely are you to use the breakdown by section/skill to understand your performance? $\alpha = .665$, indicating reasonable reliability) and 12 corresponding open items (e.g., Please explain why you have chosen this answer). Synthetic data in the forms of Figures 1 – 6 above were included as reference, and this was the first time learners had seen the six elements. The closed items were 5-point Likert scales (from 1 = Very difficult to 5 = Very easy or 1 = Definitely not to 5 = Very likely), including an ‘unsure’ option. The final open question invited further ideas or thoughts not covered in the questionnaire. The second questionnaire in week 14 referred to the lived experiences of participants using the mock exam LAD and contained eight closed items asking which of the six elements of LAD were used (again including Figures 1 – 6 above with synthetic data as a reminder and to ensure specificity in learner responses), when, and whether it helped them with their revision (RQ2) (e.g., ‘helps me create a focused revision and preparation plan for my next exam’ $\alpha = .655$, again indicating reasonable reliability). Six corresponding open items allowed participants to explain their answer and once again, a final open question invited any further ideas or thoughts not covered in the questionnaire.

Author 1, who has some insider experience as an accountancy tutor, attended full day classroom sessions and administered questionnaires. Unstructured observations from these days were collected in the form of written notes as and when the observer found behavior interesting or at odds with expectations.

2.4 Data analysis

The questionnaires were analyzed and responses were checked for normality and skewness. Descriptive statistics of the Likert scales
were created for RQ1, with the addition of Pearson correlation between PEU and PU. For RQ2, a simple count of reported use of elements was first conducted, and correlation was again used across PEU, PU and actual reported use. K-means cluster analyses were conducted based upon the pre-test PEU and PU scores of the six elements. A three-cluster solution seemed to have the most appropriate fit, whereby the final cluster solution was reached after five iterations. In respect to the free text answers, inductive and deductive thematic analysis processes were used to identify and code themes. Triangulation between the quantitative and qualitative elements of the learner questionnaires was therefore possible, informing the discussion of results and creating a rich picture of LAD use in this learning environment.

3 RESULTS

3.1 RQ1: How useful for revision do professional learners expect the Learning Analytics Dashboard (LAD) to be?

RQ1 aimed to capture professional learners’ prior perceptions of the six elements within the LAD before they were exposed to and using these dashboards. It was administered on the second day of classroom tuition (week 1), once they had been introduced to the complexities of the course and exam, but before they completed a mock exam. As previous research [18] has indicated that prior expectations might influence uptake and adoption of LAD, we specifically designed this study to capture the expectations of professional learners before they were using the LAD. Tutors were observed early and repeatedly emphasizing the absence of new knowledge acquisition in the syllabus; that passing this exam was observed early and repeatedly emphasizing the absence of new professional learners before they were using the LAD. Tutors were expected to represent the complexities of the course and exam scenario with success requiring diligent adherence to marking rules and exam technique and with completion of all requirements to the prescribed standard within the 4-hour exam time a key target. It is unsurprising then that most learners showed a positive attitude towards the LAD, recognizing its aim of explaining their mock exam results. Table 1 shows the Likert scale means (M) and standard deviations (SD) for all six elements together as well as for each element, indicating how easy each of the six elements was to use and whether the learner thought they would use them. Taking a cut-off of 3.5, 91% of participants found the six LAD elements easy to use (PEU), and 75% indicated that they were likely to use (PU) these LAD elements.

In line with TAM2, as indicated in Table 1 there was a strong correlation between PEU and PU across the overall LAD, and between PEU and PU of each of the six elements, with the exception of element 5. Subsequent analysis per element indicated substantial and significant differences in how useful participants found the respective elements, and how likely they were to use them. For example, element 6, guidance by skill (personalized feedback), was considered the easiest and most likely to use. This element had the most granular level of feedback on the LAD; written explanations of why individual marks were missed in the marking key, with some praise for marks gained. This detailed level of advice was expected to be the most useful; participant J2L05 positively viewed the “relevant and direct guidance”. Element 2, the mark breakdown by section/skill (visualization), was a close second place for a combination of easy and likely to use. This element contained the most numerical data, including marks scored as a number and percentage of marks available by requirement and skill. The data was color-coded heat map style, with bright green for higher (better) scores through light green, orange and red for lower (poorer) scores. The reasons for favoring this element were summed up in comments by participant J2L28 – “Clear labelling, categorization and signaling with colors. Can quickly identify strengths and weaknesses.”

Element 1 - your exam performance – cumulative (visualization) was considered both the hardest and least likely to use, significantly lower on both PEU and PU relative to all other LAD elements. This graphical representation of marks building towards the pass mark was intended to represent how the mark in each requirement builds towards the passing mark of 100, or 50% of available marks. However, it was the only element to record ‘Very difficult’ to understand ratings (n=4), with one participant noting they “don’t even know where to start with it” (J2L05). Another deemed it to be “rudimental” (J2L02). Even within those who indicated positively on ease of use, misunderstandings in interpretation were evident in the comments; for example, participant J1B15 gave a positive score of 5, but commented, “Easy to see per mark. Shouldn’t it be green at

<table>
<thead>
<tr>
<th>Element</th>
<th>Ease of use M (SD)</th>
<th>Likely to use M (SD)</th>
<th>Correlation between PEU and PU for each element</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average LAD score across six elements</td>
<td>3.92 (0.41)</td>
<td>3.73 (0.48)</td>
<td>.517**</td>
</tr>
<tr>
<td>1 - Your exam performance – cumulative (visualization)</td>
<td>3.04 (1.05)</td>
<td>2.90 (1.06)</td>
<td>.497**</td>
</tr>
<tr>
<td>2 - Mark breakdown by section/skill (visualization)</td>
<td>4.15 (0.71)</td>
<td>4.18 (0.61)</td>
<td>.403**</td>
</tr>
<tr>
<td>3 - What now? (static written guidance)</td>
<td>4.04 (0.68)</td>
<td>3.60 (1.03)</td>
<td>.491**</td>
</tr>
<tr>
<td>4 - Breakdown by requirement (visualization)</td>
<td>3.99 (0.91)</td>
<td>3.91 (0.80)</td>
<td>.715**</td>
</tr>
<tr>
<td>5 - Overall performance by requirement (personalized feedback)</td>
<td>4.19 (0.53)</td>
<td>3.70 (1.05)</td>
<td>.265</td>
</tr>
<tr>
<td>6 - Guidance by skill (personalized feedback)</td>
<td>4.22 (0.65)</td>
<td>4.20 (0.72)</td>
<td>.422**</td>
</tr>
</tbody>
</table>

n = 81, Pearson correlation ** p < .01
week of the program. Participants had ten days to use the LAD to
with corresponding complaints when these were not available. For
weaknesses in their performance. However, it is interesting to note
was intended, participants put value on identifying strengths and
it summarizes the performance by requirement in written text –
participants taking a more negative view argued it was “just stating
the obvious” (J1S03) and that “the table itself is more useful”
(J2L06). However, positive views of element 5 included “I like the
combination, I think one of the drawbacks of the graphs without commentary is they aren’t helping to better your performance”
(J1B04).

The thematic analysis of the open responses showed five main themes, four of which being the learners’ need for the LAD to be
clear, useful, detailed and personalized. Comments were either positive or negative around these four themes, for example, describing an element as detailed positively, not detailed enough (negative) or even too detailed (negative). No element had a universally positive or negative review from participants – even the generally low-scoring element 1 was described by one learner as a “clear visualization of performance across requirements & ES [Executive Summary – the fourth requirement] and representation of the pass mark” (J1R08). Equally, another participant thought the overall high-scoring element 2 took “too long to make sense out of” (J1L07). As expected in a learner-facing assessment LAD, and indeed as was intended, participants put value on identifying strengths and weaknesses in their performance. However, it is interesting to note that emphasis leaned heavily towards detecting weaknesses over acknowledging strong performance areas. Adding to this, the fifth theme of looking for next steps, solutions, review points, advice and explanations - preferably personalized - was identified, along with corresponding complaints when these were not available. For example, participant J2L05 praised element 6 for giving “Clear explanations catered to the answer provided” while participant J1S04 dismissed element 3 because it was “not edited based on my performance”. Our findings thus suggest that most professional learners expected the LAD would be useful to revising for their exam, and they were likely to use them. Nonetheless, substantial significant (quantitative) and subtle (qualitative) differences were noted in how useful each of the six elements were perceived, therefore it would be important to see how professionals actually used each element in RQ2.

3.2 RQ2: How did the professional learners use the LAD, and, if applicable why was it helpful to revision?
The post-use questionnaire was administered in class in the 14th
week of the program. Participants had ten days to use the LAD to
review two mock exams and were entering the intense revision
phase two weeks before the final ICAEW exam. Learners were
asked which elements they had used in reviewing mock exam
results; all had used at least one, with n = 68 (86%) learners using three or more elements and n = 32 (41%) using five elements. Figure 7 shows which elements were used, and Table 2 compares these with RQ1 perceived ease of use and likelihood of use. Whilst the least used element, element 1, mirrors learners’ expectations from RQ1, element 6 dropped to the third most used. As perhaps might be expected with soon-to-be chartered accountants, elements 2 and 4 - the two tables of marks - were, in fact, the most used elements. However, the two personalized written explanations were not far behind.

As indicated in Table 2, on average participants used 4.08 (SD = 1.24) out of 6 elements in the ten days of LAD availability, and
87% of participants (self-reported) used at least 3 elements of the LAD. In terms of help with revision items, taking a cut-off point of 3.5 again 87% found that the LAD helped them with their revision. Nonetheless, in contrast to common studies using TAM [2] PEU and PU did not significantly correlate with actual behavioral intentions (i.e., use of the 6 elements). Subsequent analyses per element (not illustrated) again showed no relations between PEU, PU and the reported use of the six respective elements. In the discussion we will explore this unexpected finding in more detail. Finally, those who were more positive about PEU and PU of the LAD were also more positive about how the 6 elements of the LAD helped with their revision.

In order to explore whether or not particular subgroups of professional learners were present in terms of prior expectations and how these subsequently might influence their LAD usage, k-means cluster analyses were conducted based upon the pre-test PEU and PU scores of the six elements. A three cluster solution seemed to have the most appropriate fit, whereby the final cluster solution was reached after five iterations, cluster 2 participants (n = 47) were the most positive about using the six LAD elements, while cluster 1 participants (n = 20) were less likely to use element 1 and element 4, while the smallest cluster 3 (n = 12) were less likely to use element 1, 2, and 5.

As indicated in Table 3, on average the three clusters used four elements of the LAD, and no significant differences were noted in
terms of overall usage. However, subsequent analyses per element did show some subtle differences in terms of usages across the three clusters.

Cluster 3 students, as expected from the pre-test, did not use element 1 at all, while 47% of cluster 1 and 22% of cluster 2 participants used element 1, which was significantly different using ANOVA. In terms of element 2, all cluster 2 participants used element 2, followed by 91% of cluster 3 participants, and 82% of cluster 2 participants, which was again significantly different using ANOVA. The subsequent elements 3-6 had no significant differences in terms of usage per cluster, whereby element 4 and element 6 were used by most participants, while element 3 was used by half of participants. What is interesting to note is the disconnect between initial expectations of cluster 1 and cluster 3 participants and whether (or not) they subsequently used these elements. For example, while cluster 1 participants indicated before that they were less convinced about element 1 and element 4, they did use element 1 the most relative to the other participants, and also 82% of cluster 1 participants used
Figure 7: RQ2 reported use of each element in the LAD (n=79)

Table 2: PEU, PU, dashboard usage, and help with revision

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Perceived ease of use PEU (pre)</td>
<td>3.93</td>
<td>0.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Perceived usefulness PU (pre)</td>
<td>3.73</td>
<td>0.48</td>
<td>.517**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Use of 6 LAD elements (0-6)</td>
<td>4.08</td>
<td>1.24</td>
<td>-.051</td>
<td>-.013</td>
<td></td>
</tr>
<tr>
<td>4. Helps with revision</td>
<td>3.93</td>
<td>0.42</td>
<td>.247*</td>
<td>.301*</td>
<td>.183</td>
</tr>
</tbody>
</table>

n = 65, Pearson Correlation * p < .05, ** p < .01

Table 3: (Self-reported) usage of six elements (in %) based upon the cluster solution of pre-test PEU and PU.

<table>
<thead>
<tr>
<th>Elements</th>
<th>1 Less likely to use element 1, 4</th>
<th>2 Most positive</th>
<th>3 Less likely to use element 1, 2, 5</th>
<th>F-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of 6 LAD elements (0-6)</td>
<td>4.24 (1.39)</td>
<td>4.11 (1.21)</td>
<td>3.82 (1.17)</td>
<td>.368</td>
</tr>
<tr>
<td>1 - Your exam performance - cumulative (visualization) (in %)</td>
<td>47</td>
<td>22</td>
<td>0</td>
<td>4.449*</td>
</tr>
<tr>
<td>2 - Mark breakdown by section/skill (visualization)</td>
<td>82</td>
<td>100</td>
<td>91</td>
<td>3.342*</td>
</tr>
<tr>
<td>3 - What now? (static written guidance)</td>
<td>53</td>
<td>53</td>
<td>55</td>
<td>0.005</td>
</tr>
<tr>
<td>4 - Breakdown by requirement (visualization)</td>
<td>82</td>
<td>89</td>
<td>91</td>
<td>0.285</td>
</tr>
<tr>
<td>5 - Overall performance by requirement (personalized feedback)</td>
<td>76</td>
<td>72</td>
<td>64</td>
<td>0.264</td>
</tr>
<tr>
<td>6 - Guidance by skill (personalized feedback)</td>
<td>82</td>
<td>75</td>
<td>82</td>
<td>0.226</td>
</tr>
<tr>
<td>Helpful with revision (1-5)</td>
<td>3.69 (.40)</td>
<td>3.98 (.33)</td>
<td>3.95 (.36)</td>
<td>4.113*</td>
</tr>
</tbody>
</table>

n = 64 (cluster 1 = 20, cluster 2 = 47, cluster 3 = 12), ANOVA * p < .05

element 4. Similarly, cluster 3 participants actively used element 2 and element 5 while they were less convinced about these elements at the start, but they remained consistent in terms of not using element 1. Finally, in terms of helpful for revision, while cluster 1 participants used the LAD elements most, relative to the other participants they were significantly less positive about how helpful these LAD elements were for their revision.

In order to unpack these quantitative findings and the different reported perspectives of the three clusters, Thematic Analysis was applied to the qualitative comments in SRQ2. Overall, learners remarked positively about the clarity and conciseness of the LAD presentation. Many of the comments were assigned to a theme of ‘identifying gaps’ – learners were keen to understand their strengths and weaknesses, with, as in SRQ1, an emphasis on weaknesses. Learners expressed a need to understand their mark, identify gaps and plan next steps. For example, participant J1L06 (cluster 2) was positive the LAD “clearly explained via heatmaps & concise steps/conclusions on my result”, whilst J2L09 (cluster 1) noted “It helps highlight ‘dark spots’ where I have a clear misunderstanding I want to ask about”, and J2L06 (cluster 3) reflected on the “thorough breakdown useful for targeting revision exam technique.” There was a distinct separation between learners who were positive about the written feedback elements – “the personalized advice is helpful for next steps” (J1B04, cluster 2) – and those who were less impressed with the automation of the personalized feedback – “Overall performance blocks were copied for R1-R3 so not helpful as same guidance repeated 3 times” (J1L06, cluster 2) and “personalized advice quite generic so not much discussion” (J1L01, cluster 1). Some were more ambivalent: “Some of the feedback felt quite generic – but some also very helpful!” (J1S05, cluster 2). Findings for our second research question therefore demonstrate high reported use of the LAD, despite patterns not being correlated...
with the intended use in RQ1. Clusters of preferred elements from PEU and PU were identified and compared to reported actual use, with some professional learners changing their mind about which of the six elements were useful in execution. Some explanation is offered in the qualitative comments, and we discuss the reasoning and implications in the following section.

4 DISCUSSION

This research was uniquely conducted in an authentic professional learning environment. Participants were professional learners, the majority aged mid to late twenties, who not only were studying for a high-stakes exam, but also working in a professional environment at post-graduate level. A key aim was to understand how and why a learner-facing assessment LAD was useful to their learning and compare this to existing LAD research conducted in the main within the more academic-focused higher education institutions. The LAD was created as a technological solution – albeit using low-level technology – to a teaching and learning problem, explaining a complex exam that qualitatively has infinite ways of being answered (both correctly and incorrectly). There are likely a multitude of these low-technology solutions across all sorts of education establishments, implemented by teachers as a practical solution to context-based problems without extensive reference to academic research on the topic. This study uses an academic lens to evaluate a practice-based solution.

In terms of RQ1, how useful for revision do professional learners expect the Learning Analytics Dashboard (LAD) to be, using TAM 91% of the professionals found the six LAD elements easy to use, and 75% indicated that they were likely to use these LAD elements. However, subsequent analysis per element indicated substantial and significant differences in how useful participants found the six LAD elements, and how likely they were to use them. This demonstrates the complexity of LAD design as highlighted by the attempt to define rules by [35]. The poor quantitative score of element 1 and associated qualitative comments evidenced the lack of understanding and misinterpretation of this visualization. Whether this can be characterized as a data-literacy problem is debatable; it is likely a combination of a less prevalent graph style (some learners recommended improvements such as more labeling or to make it vertical), the use of ‘raw’ marks instead of the more often used percentage marks in earlier ICAEW exams and the very early stage of the tuition program, when learners were still getting to grips with the marking rules.

Professional learners were also clear that a visualization being ‘easy’ did not make it ‘useful’. Element 3 in particular was criticized for being too “generic”, despite it being easy to understand. In this case, the guidance has been given in multiple other places (such as in class and in written study material), and not every learner finds the reminder helpful, despite its place as part of the mock exam results “story”. However, the mean score for PU was 3.6, above the cut off of 3.5, and so this element cannot be fully dismissed. In contrast, the strongly positive PEU and PU of element 6, personalized feedback, supports previous research espousing the benefits of feedback [19, 32] – professional learners show a strong preference for LAD to be related directly to their work. The triangulation of the quantitative results with qualitative learner comments support this insight further, with the themes of a learner inclination for LAD to be clear, useful, detailed and show next steps, summed up by participant J2L28 praising “clear labelling, categorization and signaling with colors. Can quickly identify strengths and weaknesses.” Particular emphasis was placed on identifying weaknesses (rather than strengths), supporting previous claims that LAD must support learner action [39].

In terms of RQ2, how did the professional learners use the LAD, and, if applicable why was it helpful to revision, the results indicated that on average participants used four out of six elements in the ten days that they had access to the LAD. The vast majority of participants (i.e., 87%) found that the LAD helped them with their revision for the final exam. For example, participant J1L06 was positive the LAD “clearly explained via heatmaps & concise steps/conclusions on my result.” The elements with heatmaps, elements 2 and 4, were highly used across all three clusters identified in the results analysis, supporting previous findings that traffic light style colors are useful in understanding [24]. Nonetheless, in contrast to previous research, the unexpected finding of a lack of correlation between PEU and PU and actual use could be explained firstly by the very early administration of RQ1 questionnaire. This was intended to gain the unencumbered opinion of learners, however the assessment rules of this particular exam were so unique compared to any taken previously it is possible they simply did not comprehend them well enough to infer what would and would not be useful in this context. A future study could administer this first questionnaire later in the program, when understanding is further consolidated, but before the LAD is used. Secondly, there was a high expectation of PEU and PU from element 6, the personalized feedback. Qualitative comments related to reported use demonstrate the disappointment some learners felt in this element, for example “Overall performance blocks were copied for R1-R3 so not helpful as same guidance repeated 3 times” (J1L06). It is true the sentence structure of each ‘block’ (paragraph) was the same, although the content did change depending on performance. Some learners could recognize the formula-driven nature of the feedback and considered it ‘less’ than the very specific additional feedback that some markers include on each script. For example, J1B10 wrote, “I don’t really like the formula-driven feedback, as it feels like it’s not specific and hard to pinpoint areas that the feedback is referring to sometimes.” The automated personalized feedback, whilst technically being ‘correct’ can lack the nuance of human involvement, for example being repetitive if learners were scoring in the same mark range across all requirements which was off-putting for some (but by no means all) learners.

In addition, there was criticism on the tone of the feedback from some professional learners such as “it says well done a lot & good work but only a few improvement points” (J1S02, scored 105 in the exam, only a borderline pass). This, along with the criticism of element 3 being easy but not useful in RQ1 and the emphasis on identifying weaknesses over strengths discussed above, demonstrates the difference in expectations and performance level of professional learners compared to, for example, K12 or undergraduates (such as [37]. The professional learners in this context were looking for areas to improve, they wanted to take action to do better in their next attempt. Whilst they appreciate knowing where they
were doing well, this is to know they did not need to make changes rather than a wish to be praised.

Part of the high usage was due to the LAD being embedded in the tuition program. Mock exams were not optional; they were the only way to test application of knowledge and the only way for learners to move forward. The mock exams were also an opportunity to experiment with exam techniques; to understand what works and what does not. Tutors were observed encouraging learners to try out different methods of planning, for example, in the mock exams, and the LAD showed what was and what was not successful – all this, crucially, without the need for tutor explanations. This is not to say tutors were no longer needed, but that the basic level, often asked, easily answerable from the marking key questions were addressed in the LAD, and tutor time was now available for the more complex, judgement-based questions as well as class-based discussions on content and technique.

Analysis identified clusters of learners with different preferences in their initial view of the LAD in SRQ1 and how they used the LAD in SRQ2. It seems learners were taking what they need from the LAD. It is clear from the qualitative analysis that personal preferences were not only based on the type of visualization and feedback included in the LAD, but also the story of each exam attempt. Learners expressed preferences for LAD elements which were clear and concise, and that support future work by detecting their gaps and giving guidance, as identified in previous research such as [30] – that is, to complete the cycle of learning analytics [7] with action [39]. Most learners were able to achieve this using the LAD – “Thorough breakdown useful for targeting revision exam technique” (J2L06), although some were more circumspect – “Lots of feedback – good, but I’m struggling at the moment to structure that into a plan of action” (J1S05), demonstrating that whilst the LAD is a useful tool, it cannot replace the need for learners to make linkages between actions and techniques they have implemented and the outcome as described in the LAD, neither can it replace tutor input or intervention. As others have identified [28] ‘good’ learners find the LAD useful, but those who were struggling overall might also need support to purposefully interpret and use the LAD.

This research has limitations, including the very specific context in which it is located. It refers solely to learner-facing assessment LAD and does not feature wider learning issues such as overall engagement and motivation. Instead, it assumes that a learner who had reached the point of using the assessment LAD would engage with learning and be motivated to move forward. A statistically valid relationship between LAD use and success in future exams was not identified, so the discussion focused solely on why and how learners used the LAD. There are two potential reasons for this. The first is the experimental nature of the mock exams – learners were encouraged by tutors to use them to try out exam techniques, with a resulting failure being a perfectly legitimate learning experience.

Secondly, the embedded nature of the LAD into the tuition program means recording very few participants with low reported LAD use. Due to the ethical considerations of potentially depriving learners of a useful tool, an experimental approach - in this context at least - is an unlikely practical next step. Instead, repeating the research with the next exam cohort will add to participant numbers and validity. The addition of a Self-Regulated Learning instrument to the methodology – a third questionnaire – would enable analysis to include links to this learning theory. Overall, it is hoped this research encourages further collaboration between academics and practitioners across professional learning contexts.

5 CONCLUSION

The unexpected finding of a lack of correlation between PEU and PU and actual use does not negate the high actual LAD use by professional learners in this study, and that 87% found the LAD helped them with their revision. Learners aim to pass a high-stakes exam, so they were ready and willing to use the available tools, and at this professional level of learning their focus was on understanding and improving weaknesses rather than celebrating strengths. Whilst the LAD can be described as ‘storytelling’, learners had a ‘take what you need’ approach to the available data visualizations and automated personalized feedback. Depending on both individual preferences and the exam result, not every element was immediately useful to identify weaknesses and next steps. It seems that both the ‘story’ of the mock exam and what each learner needs in the moment – facts, explanations, an overall or a detailed view – are the factors that drive LAD use. These professional learners considered the LAD as they would a menu and made their choice based on current needs. Therefore, the aim of LAD design should not be to create data displays and explanations that are liked by every learner at every moment, but that support learning at a multitude of time points and levels of success. Clarity, conciseness, understanding and guidance on next steps for improvement were considered essential by professional learners. Automated, personalized written feedback was shown to be possible and well used, although some learners recognized the automation drivers so design and written statements must be carefully created to minimize, for example, repetition. More advanced technology and AI tools may be able to help with this. Finally, the assessment LAD must be as useful to a strong learner as it is to a weak one, and offering options in a storytelling framework was convincingly effective for most in this professional learning context.

REFERENCES
