Maren Scheffel
Julien Broisin
Viktoria Pammer-Schindler
Andri Ioannou
Jan Schneider (Eds.)

LNCS 11722

Transforming Learning with Meaningful Technologies

14th European Conference
on Technology Enhanced Learning, EC-TEL 2019
Delft, The Netherlands, September 16–19, 2019, Proceedings

Springer

Mature ELLs' Perceptions Towards Automated and Peer Writing Feedback

Amna Liaqat[1(✉)], Gokce Akcayir[2], Carrie Demmans Epp[2], and Cosmin Munteanu[3]

[1] Department of Computer Science, University of Toronto, Toronto, Canada
a.liaqat@mail.utoronto.ca
[2] EdTeKLA Research Group, Department of Computing Science, University of Alberta, Edmonton, Canada
{akcayir,demmanse}@ualberta.ca
[3] Institute for Communication, Culture, Information, and Technology, University of Toronto Mississauga, Toronto, Canada
cosmin.munteanu@utoronto.ca

Abstract. Mature English Language Learners (ELLs) learning to write in informal environments have little access to instructor feedback and must rely on other sources to support their writing development. While it is known that mature ELLs trust instructor feedback, their perceptions towards feedback from non-expert sources may be mixed. We report on mature ELLs' perceptions and interpretations of peer and automated feedback when using dashboard visualizations of their writing skills derived from several metrics and sources of feedback. These perceptions and interpretations were collected through a short-term deployment of the dashboard within a writing app with 16 mature ELLs, followed by interviews with the learners. From analyses of these interviews, we suggest three design guidelines (DG) related to learning analytics dashboard design for mature ELLs in informal learning contexts. First, analytics-based feedback should contextualize ELLs' learning progress by providing temporal information about learner performance. Second, justifications should accompany feedback to avoid criticism arising from ELLs' prior beliefs. Third, learner autonomy should be fostered by offering explicit mechanisms for reflecting on feedback that is inconsistent with learner beliefs since learners are willing to question automated feedback. We discuss how these three guidelines can be used to benefit learners when an instructor is not present.

Keywords: Learning analytics · Writing · Adult learners · Migrants · Dashboards

1 Introduction

Receiving timely and meaningful feedback is crucial for writing skills development [13]. However, those studying in informal settings may not have access to instructors who can provide such feedback. One such population is mature immigrant English Language Learners (ELLs). Many of these learners are not able to receive formal education to improve their English writing skills even though they need to excel at

M. Scheffel et al. (Eds.): EC-TEL 2019, LNCS 11722, pp. 266–279, 2019.
https://doi.org/10.1007/978-3-030-29736-7_20

writing English to achieve professional and social success. Without consistent and timely feedback from instructors, mature ELLs struggle to identify errors and how to prevent them in their writing. Most of these ELLs face barriers in achieving their learning goals because of their inability to access individualized feedback tailored to them. A way to tackle this issue is to provide automated and peer feedback. However, ELLs' perceptions towards these kinds of feedback need to be explored to see the extent to which this approach can compensate for a lack of instructor feedback, as perceived expertise of the feedback source affects acceptance [22]. For this reason, this study asks "How do mature ELLs perceive and respond to automated and peer feedback on their writing?".

2 Writing Support Tools

Several tools aim to support ELL learning in informal environments. However, many tools were not designed to provide support in settings when an instructor is not present [26]. Mobile apps for vocabulary acquisition or pronunciation consist of short, spaced activities, but provide summative rather than formative feedback through gamification elements [16] or simple error counts [17]. In a study of technology use by new migrants in informal contexts, ELLs expressed that they want tools that help them plan and rehearse [12]. As well, these tools should guide them in closing knowledge gaps, especially when they are unsure of how to do so [12]. To provide such formative guidance, tools with greater socio-collaborative components are needed. Additionally, existing tools primarily focus on language skills related to vocabulary acquisition, pronunciation, and listening, rather than emphasizing writing skill development.

Several tools have been designed targeting members of other populations. These tools use peer-review processes to provide feedback for writing support: ARISE [36], Peerceptiv [32], and Peer Portal [1] are among this class of tools. Having learners assess each other's writing in this way promotes the development of evaluation and judgement skills through reflection on the peer's work, which also encourages learners to reflect on their own writing [10]. Most of these existing tools require an instructor to facilitate the peer-review activities to some degree. However, many ELLs may not be taking writing classes and therefore have no access to an instructor to manage this process. This requirement makes these systems impracticable for immigrant ELLs who need to develop their skills outside of formal learning environments. In contexts where an instructor presence is minimal, other system designs are needed for providing sustainable and meaningful feedback.

3 Learning Analytics Dashboards and Open Learner Models

Open Learner Models (OLMs) and Learning Analytics Dashboards (LADs) are feedback approaches that could be used to support this need since these student-facing-analytics can provide feedback to learners in a timely fashion without requiring instructor involvement [3]. OLMs and LADs, hereafter jointly referred to as LADs, are representations of information that a system has about a learner or group of learners [5, 7].

Traditionally, learners have been given LADs because these tools can support learner reflection and monitoring; they can even foster collaboration among peers [7].

According to the SMILI☺ framework [6, 7], several factors should be considered when designing LADs. Among these factors, is the evaluation of the tool. Ideally, field evaluations with the target users are performed to determine if the feedback is understandable. Evaluations should focus on how learners engage with the LAD, including what information they access and the accuracy of their interpretations [7]. Though many such LADs have been evaluated, they often fail to motivate design decisions and fail to analyze evaluation results with learning science concepts [29]. Few evaluations have focused on learner acceptance of the analytics, even though trust and confidence is a major barrier to learning analytics adoption [3, 19, 24]. Learner comprehension and preferences should be evaluated [38] as a first step to understanding feedback effectiveness because the objective of these tools is to motivate change. As a result, the perceived usefulness and ease-of-use should be included as part of evaluations of the potential benefits associated with LAD use [24].

The mounting body of work evaluating LADs provides evidence of their potential usefulness. However, LADs are often used within formal learning contexts, with most implementations focusing on STEM subjects [3]. These LADs may not help mature ELLs learning to write in informal environments. We, therefore, study the perceptions of this feedback mechanism by an understudied and underserved population: mature ELLs who are trying to improve their writing skills without teacher support.

4 Method

This case study was conducted to examine mature ELLs' perceptions and interpretations of automated and peer feedback delivered via an LAD from a user-centered design perspective [31]. The LAD implements visualizations of learners' writing skills, derived from several automated metrics and sources of feedback (expert and peer). We collected writing samples from immigrant ELLs through a short-term deployment of an app that provides both automated and peer feedback. We then conducted post-deployment focus groups where participants were presented with dashboards to gauge their perceptions of automated and peer feedback.

4.1 Participants

The study was approved by the university's research ethics board. The first author visited classes in the Language Instruction for Newcomers to Canada (LINC) program in a large, predominantly English-speaking metropolitan area to invite students to participate. LINC is a government-funded program offering free English-language classes to recent migrants: 16 mature ELLs (Female = 13) consented and received an honorarium of $50 and reimbursement for travel expenses to the study site.

The gender split in this study is representative of that found in LINC classes (72% of students are female [18]), where students are assessed using the Canadian Language Benchmarks (CLB) standard before placing them in classes. The CLB is a scale describing language proficiency that has three stages. At the time of the study, all participants were CLB stage 2. Individuals in the second stage can participate in a variety of contexts and independently engage in routine and familiar situations [20].

The average age of participants was 38.56 (SD = 6.48). Seven participants spoke Farsi, five spoke Mandarin/Cantonese, and each of the remaining participants spoke one of French, Italian, Spanish, Ukrainian, Russian, Korean, Portuguese, or Azari. Excluding English, four of the participants spoke more than one language. All participants held at least a college diploma or a bachelor's degree. Additionally, some had a master's degree (six) or a PhD (two).

This sample reflects the Canadian immigration system, which favours highly-educated immigrants selected using a competitive point-based system. Most participants (11) were unemployed at the time of the study. Three worked part time and two had full-time work. For eight participants, improving English for daily life was an important motivator for taking English-language classes, followed by getting a job (three), preparing to study (two), passing a test to get certified in a trade or profession (two) and preparing for a citizenship test (one).

4.2 Dashboard Data Sources

In this section, we introduce each measurement used to create the dashboard that provides automated and peer feedback. These measurements include feedback from an instructor using a rubric and an automatically generated score.

CELPIP Derived Rubric. An independent instructor who specializes in adult ELL instruction derived an assignment grading rubric based on the Canadian English Language Proficiency Index Program (CELPIP) [37]. CELPIP is a standardized exam that measures the test-taker's communication abilities in informal, routine contexts, such as interacting with coworkers and friends. The CELPIP was selected for this study as it is a standardized rubric for ELL informal writing, which was the focus of this study. The rubric consists of four dimensions: Task completion and coherence, format and tone, mechanical convention, and lexical resource. These dimensions are scored on a scale of 1 (some proficiency) to 5 (advanced proficiency).

Instructor Feedback. The same instructor who created the rubric used it to grade all the assignments submitted during the deployment. The instructor provided scores for each of the four dimensions. The instructor also wrote a brief (three to six sentences) profile for each learner based on all the assignments (up to three) submitted by that learner. The instructor was asked to base the content of the profiles on the rubric and to include observations of the learners' writing strengths and weaknesses, to provide direction for improvement, and to comment on any general trends across assignments. This profile was the instructor feedback that we compared learner reflections against.

Automated Scoring. In the dashboard, learners were presented with automated scores for each of their assignments. Assignment scores were predicted by running simple linear regression on the first assignment (n = 14), which had been graded by the instructor to generate equations for predicting instructor scores. Feature selection was done with SiNLP (Simple Natural Language Processing Tool). SiNLP is a linguistic analysis tool that evaluates 17 features of writing (e.g., number of pronouns and number of future words). SiNLP was used for feature analysis because it is a simple tool that has been shown to provide similar levels of accuracy to more complex discourse analysis tools, such as Coh-Metrix, when predicting essay scores [11].

From the feature set produced by SiNLP, a subset was selected using WEKA's CfsSubsetEval method with best first search to identify the features that most accurately predicted instructor scores. WEKA is a software package that provides tools for data analysis and predictive modelling. The CfsSubsetEval method evaluates the subset of features with the highest predictive power while minimizing inter-correlation [21]. Feature selection was done for the overall score, as well as for each of the four dimensions. Next, simple linear regression was run on each of the five scores. The resulting equations were used to calculate the predicted scores for the three assignments (Eqs. 1–5). Definitions for the features are taken from [11] and provided below:

- TTR (Type Token Ratio): A measure of lexical diversity computed by dividing the number of types (categories of words) in the text by the number of tokens (total words) in the text, with a higher value indicating more diverse vocabulary use.
- F (Future): A measure of text temporality. Tense use can indicate the rhetorical stance and cohesion of a text.
- NW (Number words): The total word count of the text. Text length is related to discourse sophistication and structure.
- SPP (Second person pronouns): This count can be used as a measure of anaphor use (referencing earlier parts of the text) and can indicate text coherence.
- N (Negations): A count of a type of connective that indicates a contradiction (e.g., "however", "but"), and it is a measure of text coherence.
- D (Demonstratives): A count of words such as "this", "that", and "these". Demonstratives indicate references to information present elsewhere in the text, and they serve as a measure of cohesion.

$$\text{Task Completion and Coherence} = 11.9\,\text{TTR} + 25.2\text{F} + -5.0 \tag{1}$$

$$\text{Format and Tone} = -0.0073\text{NW} + 13.8\text{SPP} + 107.6\text{N} + 34.3\text{F} + 3.2 \tag{2}$$

$$\text{Mechanical Conventions} = -19.3014\text{D} + 3.6 \tag{3}$$

$$\text{Vocabulary} = 5.1\text{TTR} + 23.8\text{F} + -0.5 \tag{4}$$

$$\text{Total} = 5.5\text{TTR} + -15.8\text{D} + 21.7\text{F} + -0.4 \tag{5}$$

The predicted scores resulting from the equations were compared with the instructor graded scores. The predicted scores from our equations were fairly accurate. Across all four dimensions and three assignments, there was an average difference of 0.62 points (SD = 0.50) on a 5-point scale between the predicted and instructor-assigned score.

4.3 The Dashboard Visualization

The visualization (Fig. 1) was designed to display a line graph of a user's automated score for each of the three assignments in blue. The red line displays the average score across all participants. This provides learners with a temporal view of their performance, as was suggested by [15, 25]. Scores were rounded to nearest the .25 point. This accounts for a portion of the uncertainty associated with automated scoring, as suggested by [14]. If an assignment was not submitted, the score was displayed as zero.

Each user receives five graphs: one for each of the four dimensions and one for the overall score. Below each graph, general feedback is provided via text. This feedback is drawn from the rubric feedback corresponding to the average score across the three assignments. This general feedback uses text to provide further context to the graph above, as suggested by [2]. The general feedback is intended to provide users with a holistic impression of their performance. Below the general feedback, users can view the peer-feedback they have received for each assignment. This was intended to provide detail and clarity to the general, automated feedback. Per the framework proposed by [6], the analytics were designed to allow learners access to different levels of detail. Learners could view a general overview or explore a single dimension. Within each dimension, learners were provided with a temporal view of how they performed across the three assignments and could access peer-feedback for each assignment.

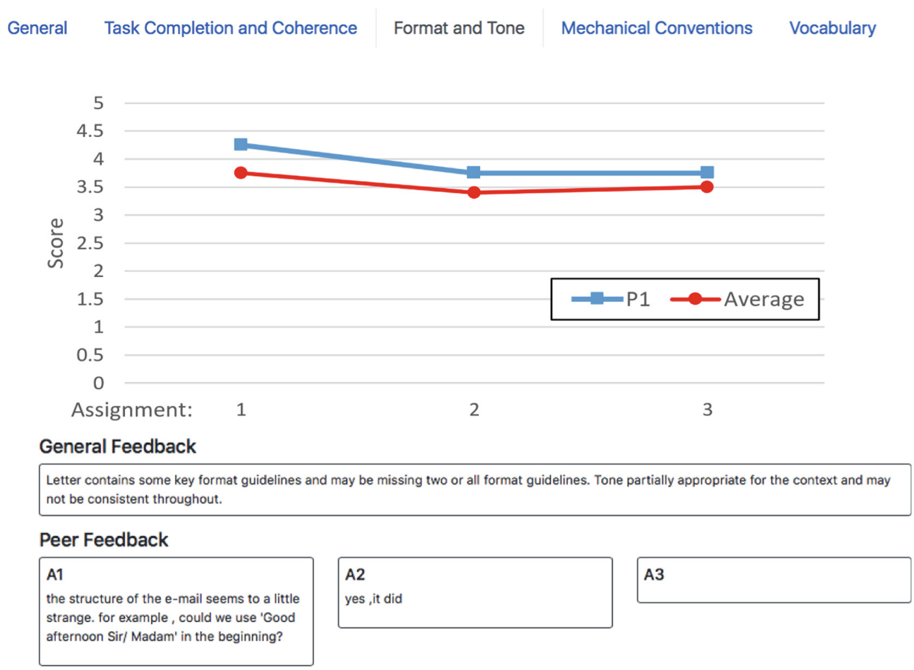

Fig. 1. The visualization of the dashboard combines automated and peer feedback to present data on performance over time, across dimensions, and on each activity. (Color figure online)

4.4 Study Procedures and Learning Tasks

Participants completed three informal writing assignments and used the app to provide peer-feedback over nine days, with an assignment due every three days. To submit assignments as well as provide and receive feedback, participants used a web-app that runs in any browser but is more suitable for larger screens (e.g., laptops). In the app, participants receive writing prompts and can submit a response. Users are also assigned a partner. After the participant submits a writing assignment, the partner can provide peer-feedback and vice versa. During peer-review, reviewers are asked four general questions to guide their feedback. Each question corresponds to a dimension on the CELPIP rubric and is listed below:

1. Did the letter address all the main points required to complete the task? Which parts of the task are missing? (Task completion and coherence)
2. Was the letter organized well so that it was easy to understand? What can be done to ensure good flow and organization? (format and tone)
3. Did the ideas of the writer connect well? How can this be improved? (mechanical convention)
4. Did the writer use a wide range of vocabulary for the task? How can this be improved? (lexical resource)

After the activities were complete, all participants were invited to the lab for a focus group session where they were presented with the dashboard. Twelve of the 16 participants attended the group session phase. The four participants who were unable to attend a group session (due to scheduling constraints) met with the researcher one-on-one online. The 12 participants attending the focus group sessions each joined one group, for a total of three groups, consisting of three, four, and five mature ELLs. As almost all participants attended ELL classes at the same centre, participants in the focus groups generally knew each other. All three focus groups were audio recorded and transcribed, and they were used to provide context when answering our research question.

Author 1 first demonstrated the visualization. Participants were informed that the information they received was not produced by an expert and may be inaccurate. They were given time to interact with their visualizations and reflect on the following questions: "Based on what you see, what would you say are your writing challenges? What are you good at? How do you think you can improve? Please write a few lines reflecting on your observations."

After everyone had submitted their responses, the researcher led the group through several questions about their interaction with and perceptions of the dashboard. As all participants were at an intermediate to advanced English proficiency level, they were able to understand and participate in the discussion. The below questions were asked:

1. Did you find the automated scores accurate? The peer feedback? Why or why not?
2. Do you think the feedback (scores, general, and peer) was helpful?
3. Did seeing the average scores of all the learners help you understand anything about your own performance?
4. Did you feel surprised or anxious about any of the information your received?

4.5 Data Analysis

Data gathered from students' written reflections and the focus-group transcriptions were analyzed using inductive data analysis [28]. In this approach, there are no pre-developed schemes or templates: codes emerge from the data. This coding procedure was part of an overarching analysis approach where the flow model of content analysis was used [30]. This model employs three steps in sequence: (1) data reduction, (2) data displays, and (3) conclusion drawing and verification. In Step 1, data was reviewed to determine patterns and codes, independent of the type of code displayed. Step 2 included reorganizing data to make the patterns and codes more explicit and easily accessible. In Step 3, themes were grounded in the data and clearly appeared from those suggested in Step 1 and Step 2. Finally, verification was performed by repeating all of the steps three times.

The peer-feedback data was not appropriate for content analysis because it is limited both in number and content (e.g., "Yes", "yes, it did"). While we could not reliably analyze the feedback provided by peers as a result of these limitations, we report participants' opinions about peer feedback. These data came from the focus groups.

For data analysis, each participant's reflection was compared against the information presented in the LAD. Through this process, it was determined how accurate learner reflections on their writing skills and learning process were (i.e., how closely their perceptions aligned with the dashboard). Next, the strengths and weaknesses identified in participants' reflections were compared with those contained in the instructor's feedback.

Analyses were handled by a researcher (Author 2) who is experienced in qualitative data analysis. Additionally, another expert in the field (Author 3) reviewed all of the steps of this analysis and confirmed the output.

5 Findings

We report our findings in accordance with the themes that emerged during data analysis. These themes consist of learners' focus on challenges over strengths, evaluation of performance over time, incorrect interpretations possibly tied to past beliefs, and a tendency to question automated and peer feedback.

5.1 Focus on Challenges Over Strengths

Participants stated 11 strengths and 26 weaknesses. Almost half the students (n = 7) only specified their weaknesses without discussing any of their strengths. These findings suggest learners were focused on identifying weaknesses rather than strengths in their writing. This tendency towards understanding weaknesses to improve their writing skills also can be seen in nearly all participants' (n = 15, 93.8%) expressed desire to improve further. Learner identified methods for improving their writing skills usually centered on practicing more (n = 9). Other approaches included finding sources of additional feedback and guidance, as stated by P5, "the key is to have some professors to review and to give advice", and investing more time, as was stated by P9: "I need really do more practice and more time to improve my level". These expressions may also be evidence of participants' high motivation, which would be consistent with prior work showing that mature ELLs have high intrinsic motivation for learning to write [32].

5.2 Evaluation of Performance Over Time

Almost half of the ELLs (n = 7, 46.7%) reviewed their performance by looking at their improvements throughout the app deployment, as can be seen through P13's comment that "My general feedback about mechanical convention was near to average and was progressive in my third assignment", and P9's comments that "In first practice in task completion and coherence I was lower than average but after I understood my weak points so I arrived near the average point and the same thing happens for format and tone parts."

This behaviour is consistent with that of other adult language learners who have used this class of feedback tools [13, 15]. However, this type of comparison goal is not typically supported within the visualizations we provide to learners [14] as temporal analytics are a relatively new area of exploration [25].

5.3 Incorrect Interpretations Possibly Tied to Past Beliefs

Participant interpretations of their feedback contain incorrect or sub-optimal interpretations of both their strengths (n = 5, 45.5%) and weaknesses (n = 8, 30.8%). While the percentage of potential misinterpretation of weaknesses is almost double that of their strengths, this rate is consistent with the rate at which they identified strengths and weaknesses. An example of an incorrect interpretation from P14 (Fig. 2) demonstrates how participants interpreted the visualizations. P14 stated "My challenges are the mechanism [mechanical conventions] and vocabulary". However, Fig. 2 shows the participant's performance with respect to mechanical conventions was above average for the two assignments he had submitted; the third assignment measure is missing because it was not submitted. The visualizations for the task completion and coherence dimension (Fig. 2) indicate P14 has an area where he is weaker, which this learner failed to see. This makes P14's identification of the mechanical conventions dimension as his primary weakness an incorrect interpretation of the provided feedback.

Participants' suboptimal or incorrect interpretations may come from their past experiences and the prior beliefs that stem from these experiences [33]. These prior beliefs likely play a role in mature ELLs' interpretation of these charts since we know that members of this population can possess strong epistemic writing beliefs [27].

5.4 Tendency to Question Automated and Peer Feedback

Average scores shown in the visualization were perceived as helpful but providing more details would have improved perceived usefulness: "I see the line and it makes sense but [inaudible] the structure it's very weak for my writing I need to improve more and more, I don't know. Just from the scores, it's maybe not enough" (P5). P5 added there is "not so much context" and suggested model assignments for different scores:

> In my opinion not just the score. It shows that you have some gaps from others, so I need to improve to make scores improve. It just shows the scores. If possible, it could show some model assignments to show us how others write. (P5)

The request for more detail was agreed to by P2 ("yeah") and P4 ("Yeah, I agree").

Participants think that all feedback (scores, general, and peer) was useful, in general. However, the peer feedback was perceived as unreliable because it was not always available: "my partner didn't respond to me for the second assignment. So, I think that affected my feedback and my graph is strange" (P13). These perceptions carried over to the writing platform with most participant opinions focusing on how helpful or "very useful" (P16) it was for them. Comments included:

> I can see others people's writings, and it helps me a lot. But maybe it should provide more partners at the same time. Because one partner's writing skills are not enough, sometimes she couldn't give me the correct advice. After all, I like this program (P7)

> Using the App has helped me to have a better understanding of what I was asked for. In short, I could say it's been a good practice. (P16)

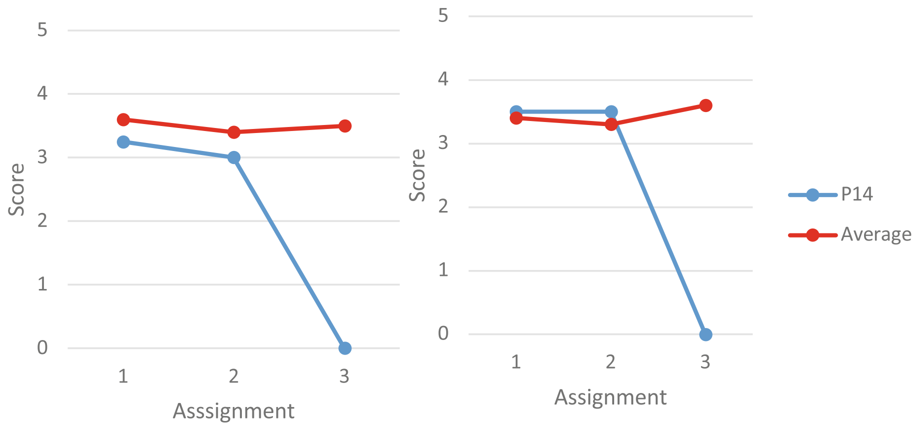

Fig. 2. P14 performed below average in the task completion and coherence dimension (left). However, P14 identified mechanical conventions as a weakness despite better performance, which was above average, in this dimension (right). (Color figure online)

5.5 Design Guidelines for Feedback in Informal Learning

In light of the above findings, we suggest three design guidelines (DG) to consider when creating feedback tools for mature ELLs in informal learning contexts.

DG1: Feedback should contextualize performance by showing how learners are progressing over time, while allowing learners to compare their performance against a reference point.

DG1 is based on our finding that when presented with two dimensions of comparison (temporal versus peers) mature ELLs chose to evaluate progress by looking at their performance over time. Temporal analyses that present learners with historic data on past performance can prompt reflection on performance over time [15]. While these participants did not strongly emphasize comparison with peers, it is a common reference point used to contextualize analytics of student performance [23, 35], suggesting it could be used in this context. Moreover, some participants compared the average score to their own when identifying gaps in their performance and through that their need to seek strategies for improvement. Participants also requested access to sample assignments to make sense of expectations. If peer work is used to provide these exemplars, it would give learners the opportunity to learn from stronger peers [8]. Therefore, it may be beneficial to allow mature ELLs the choice to view peer scores and samples.

DG2: Feedback should be presented with clear and detailed justification to prevent possible bias arising from mature ELLs' prior beliefs.

Our mature ELLs already have substantial educational experience and possess a strong skillset for achieving learning success, suggesting that we need to design learning activities and tools that recognize and support this learner characteristic. Along with this prior experience, our findings suggest mature ELLs hold pre-existing beliefs about their writing strengths and weaknesses. This is indicated by learner reflections where they identified weaknesses that were not included in the dashboard or that contradicted the information presented there. One contributing factor to their strong writing beliefs may be that our participants have completed post-secondary degrees and have likely acquired learning skills and beliefs they are comfortable with. Therefore, presenting mature ELLs with information on their performance may not be sufficient. While some groups of learners can benefit from receiving summarized performance reports (e.g., lower achieving students) [6], mature ELLs may benefit from access to their full, detailed student models. As experienced learners have well established beliefs, they may interpret the information in a manner that confirms those beliefs. Thus, in informal learning with mature ELLs, consideration should be given to helping learners identify when their beliefs are incompatible with their skills or performance so that the system can scaffold belief revision.

DG3: Foster learner critical thinking and autonomy using mechanisms that support learners' tendency to question automated feedback.

We found that mature ELLs were comfortable questioning scores they disagreed with. This may be because learners do not perceive automated feedback as having the same authority as that provided by an instructor. Perceiving automated feedback as having less authority may benefit learners because those who view the teacher's role as one of authority take less responsibility for their learning [9]. Additionally, online platforms in blended language-learning classes have been shown to increase learner awareness of feedback importance, improve confidence, and trigger a shift in learner perceptions of the instructors' role from that of director to that of facilitator [34].

Thus, we find automated feedback could play an important role in scaffolding learners towards critical assessment of their writing by offering explicit mechanisms for users to challenge the feedback or to reflect on why they may disagree with it, as is commonly done in negotiated and persuadable open learner models [4].

6 Limitations

Our participants consisted of a specific subset of ELLs (highly educated), thus our findings may not be representative of other immigrant contexts. In our analysis, we were unable to include peer-feedback as it lacked detail or was not provided. So, the role of peer-feedback in prompting learner reflections has not been assessed. In future studies, mechanisms should be designed to elicit more detailed, meaningful peer feedback.

7 Conclusion

In this paper, we explored mature ELLs' perceptions of writing skills visualizations, derived from several automated metrics and sources of feedback (expert and peer). The importance of providing such types of feedback comes from the lack of available instructor feedback for our target population, immigrants. This population usually does not have access to formal language education, even though their language proficiency is one of the biggest factors affecting socio-economic status in their new country. In this sense, a dashboard that provides customized feedback for the writing activities they perform on their own contributes not only to the success of individuals but also the development of community. Based on our findings, we presented three design guidelines that can be used to help others create similar types of systems within their contexts.

Future studies should employ long-term deployments and explore ways to facilitate high quality peer feedback. A study exploring the effectiveness of peer and automated feedback compared to instructor feedback could show whether these practices influence language learning. Alternatively, a similar type of technology could be built to support the development of recent migrants' speaking skills with automated and peer feedback of learner speech being used to advance their fluency and pronunciation accuracy.

Acknowledgements. This work was supported by AGE-WELL NCE Inc., a member of the Government of Canada's Networks of Centres of Excellence.

References

1. Aghaee, N., Hansson, H.: Peer portal: quality enhancement in thesis writing using self-managed peer review on a mass scale. Int. Rev. Res. Open Distrib. Learn. **14**(1), 186 (2013)
2. Ali, L., Hatala, M., Gašević, D., Jovanović, J.: A qualitative evaluation of evolution of a learning analytics tool. Comput. Educ. 58(1), 470–489 (2012)

3. Bodily, R., et al.: Open learner models and learning analytics dashboards: a systematic review. In: Proceedings of the 8th International Conference on Learning Analytics and Knowledge, pp. 41–50 (2018)
4. Bull, S., Ginon, B., Boscolo, C., Johnson, M.: Introduction of learning visualisations and metacognitive support in a persuadable open learner model, pp. 30–39, April (2016)
5. Bull, S., Kay, J.: Open learner models. In: Nkambou, R., Bourdeau, J., Mizoguchi, R. (eds.) Advances in Intelligent Tutoring Systems, pp. 301–322. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-14363-2_15
6. Bull, S., Kay, J.: SMILI☺: a framework for interfaces to learning data in open learner models, learning analytics and related fields. Int. J. Artif. Intell. Educ. **26**(1), 293–331 (2016)
7. Bull, S., Kay, J.: Student models that invite the learner in: the SMILI:() open learner modelling framework. Int. J. Artif. Intell. Ed. **17**(2), 89–120 (2007)
8. Bull, S., Nghiem, T.: Helping learners to understand themselves with a learner model open to students, peers and instructors, April 2003
9. Ceylan, N.O.: Fostering learner autonomy. Procedia Soc. Behav. Sci. **199**, 85–93 (2015). https://doi.org/10.1016/j.sbspro.2015.07.491
10. Cho, Y.H., Cho, K.: Peer reviewers learn from giving comments. Instr. Sci. **39**(5), 629–643 (2011). https://doi.org/10.1007/s11251-010-9146-1
11. Crossley, S.A., Allen, L.K., Kyle, K., McNamara, D.S.: Analyzing discourse processing using a simple natural language processing tool. Discourse Process. **51**(5–6), 511–534 (2014). https://doi.org/10.1080/0163853X.2014.910723
12. Demmans Epp, C.: Migrants and mobile technology use: gaps in the support provided by current tools. J. Interact. Media Educ. **2017**(1), 2 (2017)
13. Demmans Epp, C.: Protutor : a pronunciation tutor that uses historic open learner models. University of Saskatchewan (2010)
14. Demmans Epp, C., Bull, S.: Uncertainty representation in visualizations of learning analytics for learners: current approaches and opportunities. IEEE Trans. Learn. Technol. **8**(3), 242–260 (2015). https://doi.org/10.1109/TLT.2015.2411604
15. Demmans Epp, C., McCalla, G.: ProTutor: historic open learner models for pronunciation tutoring. Artif. Intell. Educ. **2011**, 441–443 (2011)
16. Edge, D., Cheng, K.-Y., Whitney, M., Qian, Y., Yan, Z., Soong, F.: Tip tap tones: mobile microtraining of mandarin sounds, pp. 427–430, September 2012
17. Edge, D., Searle, E., Chiu, K., Zhao, J., Landay, J.A.: MicroMandarin: mobile language learning in context. In: Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems - CHI 2011, Vancouver, BC, Canada, 2011, p. 3169 (2011)
18. Evaluation of the Language Instruction for Newcomers to Canada (LINC) Program (2011). https://www.canada.ca/en/immigration-refugees-citizenship/corporate/reports-statistics/evaluations/language-instruction-newcomers-canada-2010/intro.html#a2
19. Greller, W., Drachsler, H.: Translating learning into numbers: a generic framework for learning analytics. J. Educ. Technol. Soc. Palmerston North **15**(3), 42–57 (2012)
20. Hajer, A., Kaskens, A.-M.: Canadian Language Benchmarks: English as a Second Language for Adults. Citizenship and Immigration Canada (2012)
21. Hall, M.A.: Correlation-based feature selection of discrete and numeric class machine learning. University of Waikato, Department of Computer Science (2000)
22. Ilgen, D.R., Fisher, C.D., Taylor, M.S.: Consequences of individual feedback on behavior in organizations. J. Appl. Psychol. **64**(4), 349–371 (1979)
23. Jivet, I., Scheffel, M., Drachsler, H., Specht, M.: Awareness is not enough: pitfalls of learning analytics dashboards in the educational practice. Data Driven Approaches Digital Educ. **2017**, 82–96 (2017)

24. Jivet, I., Scheffel, M., Specht, M., Drachsler, H.: License to evaluate: preparing learning analytics dashboards for educational practice. In: Proceedings of the 8th International Conference on Learning Analytics and Knowledge, pp. 31–40 (2018)
25. Knight, S., Wise, A.F., Chen, B.: Time for change: why learning analytics needs temporal analysis. J. Learn. Anal. 4(3), 7–17 (2017)
26. Kukulska-Hulme, A., Shield, L.: An overview of mobile assisted language learning: from content delivery to supported collaboration and interaction. ReCALL **20**(3), 271–289 (2008). https://doi.org/10.1017/S0958344008000335
27. Liaqat, A., Munteanu, C.: Towards a writing analytics framework for adult english language learners. In: Proceedings of the Seventh International Learning Analytics & Knowledge Conference (LAK 2017), Sydney, Australia (2018)
28. Mackey, A., Gass, S.: Second Language Research, Methodology and Design. Lawrence Erlbaum Associates (2005)
29. Marzouk, Z., et al.: What if learning analytics were based on learning science? Australas. J. Educ. Technol. **32**(6), 1–18 (2016)
30. Miles, M.B., Huberman, A.M.: Qualitative Data Analysis: An Expanded Sourcebook, 2nd edn. Sage Publications Inc., Thousand Oaks (1994)
31. Norman, D.A.: User Centered System Design: New Perspectives on Human-Computer Interaction. CRC Press, Boca Raton (1986)
32. Peerceptiv - Data Driven Peer Assessment. http://www.peerceptiv.com/wordpress/
33. Ross, S.: Self-assessment in second language testing: a meta-analysis and analysis of experiential factors. Lang. Test. **15**(1), 1–20 (1998)
34. Snodin, N.S.: The effects of blended learning with a CMS on the development of autonomous learning: a case study of different degrees of autonomy achieved by individual learners. Comput. Educ. **61**, 209–216 (2013)
35. Verbert, K., Duval, E., Klerkx, J., Govaerts, S., Santos, J.: Learning analytics dashboard applications. Am. Behav. Sci. **57**(10), 1500–1509 (2013)
36. Vista, A., Care, E., Griffin, P.: A new approach towards marking large-scale complex assessments: developing a distributed marking system that uses an automatically scaffolding and rubric-targeted interface for guided peer-review. Assessing Writ. **24**, 1–15 (2015). https://doi.org/10.1016/j.asw.2014.11.001
37. What is CELPIP? https://www.celpip.ca/what-is-celpip/. Accessed 10 Jan 2018
38. Zapata-Rivera, D., et al.: Designing and evaluating reporting systems in the context of new assessments. Augmented Cogn. Users Contexts **2018**, 143–153 (2018)