Reviewer 1 The rationale of the study has been adequately contextualized with respect to previous research, especially in the fields of 'emergency remote teaching' (ERT) and reflection practices applied in educational contexts. However, the reflective skills that are expected to be promoted within an L2 learning environment could also be related to the promotion of 21st century skills, which are mentioned at the end of this section (lines 180-181), because this connection could further justify the importance of this research. Reference to 21st skills could potentially be made earlier in this section, provided it is supported by relevant in-text citations, which are currently missing.	Thank you reviewer 1 for the well-structured and well directed practical feedback recommendations for me to execute. Line 85 added Reflection is a transferrable competence in lifelong learning where learners feel in control over future actions. Reflective learning is a 21st century skill that grows awareness to help adapt to the challenges of technology and ever-changing environments (Colomer et al. 2020) In references Colomer, Jordi, Teresa Serra, Dolors Cañabate, and Remigijus Bubnys. "Reflective learning in higher education: Active methodologies for transformative practices." (2020): 3827.
Regarding the use of sources in this section, it seems that in line 80 the source mentioned should be further elaborated, because it is not very clear how the aim of the present study relates to the research by Vogt et al (2020). This addition could also strengthen the argument made in this paragraph.	in order to understand how the innovations to assessment were received by the students new to working in a remote context which would be beneficial for course designers and language instructors. Language students with no previous experience of ERT language learning, offer perspectives on experiences of e-assessment and what they considered conducive to learning which builds on a European study of English learners and their experiences of assessment (Vogt et al. 2020).
Introduction and Results In relation to the research questions (lines 94-96), the second research question could be further specified (e.g. What was learned… in terms of …?). Moreover, the third question seems a bit vague and could be further specified as well.	• 1 What did students report on learning about the language and culture content from completing assessments? • 2 What qualities of the e-assessments did the students value as they executed the tasks? • 3 How did students orient towards the e-assessment tasks in the reflection?

in the 'Results' section there is not a clear correspondence between the way the findings are presented and how they answer the research questions posed in the 'Introduction'. This could be improved if the 'Results' section included a different grouping of the transcribed extracts, so that the answers to the research questions were more direct and clear. Alternatively, these answers could be added as comments in the 'Discussion' section.	I placed the question above the relevant section and reordered the presentation of results to match table 4.
Introduction and Methodology The option of transferring the information presented in sub-sections 1.2. and 1.3 to Section 2 (Materials and Methods) should be considered, because the information about the research context (setting, established teaching and assessment practices), as well as the description of the course assessment tasks used as part of the research implementation, typically constitute elements of the research method.	Done!
Line 174: Was the task marking rubric (appendix 2) based on specific criteria for reflective practice set in previous studies? The addition of relevant research would strengthen the validity of the marking rubric.	Yes included The rubric was designed by the teacher/researcher as a tool for the students to production the reflection and for the assessor to award marks. Just as rubrics guide students on assessor expectations in other assessments, Chen and Chan (2019) experimented by incorporating a rubric as a resource for reflection. They found the rubric promoted students' level of reflection. The rubric was written in simple clear language based on issues reported on student use and understanding of reflection language as reported in the Asina et al.'s (2018) Narrative Assessment Reflection Rubric validation study in higher education.
2.Materials and Methods **2.3. Data collection** Line 218: It is not very clear how some submissions were considered as 'partly irrelevant'.	I have provided an example. It is long and if the editor is able to keep it , it shows what was coded as other because the reflection was off task.
Table 2 and lines 221-229: The comments made under Table 2 should follow the order in which the information is presented on the table, so that clarity can be increased. In particular, the multimodal task types could be presented on the table in the same order as the comments made under the table. Alternatively, *the task types could be presented on the table in order of frequency* (e.g.higher to lower). Moreover, Table 2 should follow (and not precede) all the relevant comments made in the text.	Table follows the paragraph

In lines 221-226, the information could be rephrased, in order for more emphasis to be placed on the types of tasks that gathered a higher percentage. The division of this part into two separate sentences could be tried.	I moved it to course assessment tasks as reviewer 3 suggested. I split in two and added reference about non threatening environment because the practice 5 and choose the best for marking	
Lines 226-227: unclear information and phrasing. To which multimodal task (Table 2) do the 'summative language and culture final' correspond to?	I entered final in the tables to show which one it corresponds to	
Lines 227-229: this comment should be rephrased to sound more formal. Another option would be for this observation to be mentioned in the 'Discussion' section only.	I removed it reviewer 3 said no room for ' I ' sentences	
2.4. Analysis of the transcribed oral reflection tasks This part of the methodology should include some references to similar research, in which analysis of qualitative data had been conducted. Relevant sources should also be used to support the concepts mentioned in line 243.	Methodology reference Green, Alison. 1998. *Verbal Protocol analysis in language testing research: A handbook* (Vol. 5). Cambridge: CUP Similar research Ducasse and Brown (2009) , Brown and Ducasse (2019	
3.Results Line 266: the "three overarching categories" should be clearly named, maybe within parentheses.	Three overarching independent categories could be easily and reliably identified, each concerned with a discrete focus of the reflection and together accounted for almost all the data. These comprised Learning, Doing assessment and **Reflecting on learning and doing assessment.** Table 3 lists the reflection focus categories with an expansion into sub-themes that arose from further analysis within the main categories. The 'learning' theme divides into comments on the content and the language. The 'doing assessment' theme is broadly defined by all things related to the task and has no subtheme. The reflecting on learning and doing assessment theme divided into two overarching themes, i.e., the personal and affective, and the interpersonal. The personal and affective sub-theme subsumed a large number of examples to define it including self-awareness, affective responses and appraisal.	
Lines 302-303: the information should be rephrased for more clarity.	Line 302 corrected The personal / interpersonal	
General comment for the 'Results' section: Overall, it should be noted that this section should directly answer the research questions mentioned in the 'Introduction' (lines 94-96).	I pasted in the RQs per section	

An alternative way of organizing the 'Results' section would be the following: the content of the students' reflection could be analyzed in two basic dimensions "personal" and 'interpersonal" in relation to the three overarching categories.	
The rationale behind the division of the three overarching categories into sub-themes is not very clear. The merging of some sub-categories that present common characteristics could be considered (e.g. subtheme 3.1.3.2 could be merged with 3.1.3), in order for clarity and comprehensibility to be increased. Moreover, there should be more analysis on some transcribed extracts, whose length could probably be shortened. For example, in subtheme 3.1.1. (Affective response) there could be some comments on how exactly the students felt about specific tasks, that is, which task was "more challenging", "difficult", "interesting", "real" and how these perceptions related to their feelings of "confidence", "enjoyment" or "distress". In the same way, in 3.2 (Theme 2), there should be more specific comments on the challenges the students faced and the skills they had the chance to practise. More specific comments are also needed in subthemes 3.3.1.1. and 3.3.2.	I merged some categories as advised Included: The Video, Reading and writing report was challenging for students because they were required to apply the academic skill of synthesis in Spanish (extract 22). The culture essay was enjoyed because the topic was current or 'real' as expressed in extract 24. For the student it made it interesting to write and gave them a ' good feeling' to be writing so proficiently on an area of interest related to their studies (food sustainability and the environment). They were aware of weaknesses e.g., extract 27, on copying too much text in a summary task and when students were assessed on their weaker skills such as listening, they felt insecure (extract 29). Developing proficiency was also recognized (extract 26) where a heritage speaker acknowledges previous limitations in range and extract 28 where they feel competent and confident after a presentation and are prepared to speak more Spanish.
Discussion – Conclusion There might be some additional analysis regarding the wider implications of this study in the field of alternative assessment practices in L2 instruction, which could probably take place within both live and online teaching environments.	The wider implications that this study brings to the field, is the implementation of alternative assessment practices in L2 instruction, which could take place

Implications could also be considered in relation to student-centeredness and the exploration of adult learners' needs, as a means of facilitating the L2 learning process in emergency situations	within F2F or online teaching environments. This exploration in the ERT context provides and insight into student centred learning occurring in bedrooms, studies, gardens and loungerooms globally. Learners worked without the support of buddies or classmates so analysing the content of their reflections on achievements and future need to improve, was a unique way of gaining insight into L2 adult language learning needs and processes in educational emergency situations.
Text formality: Some revisions could probably be made in an effort to reduce the use of the personal pronoun 'I", especially in the 'Discussion' section.	Yes.. In the discussion, as teacher /researcher, personal reflections are offered on the findings from this explorative study and how they could be seen to provide evidence that the students valued e-assessments and attributed learning content, language and personal skills through the tasks. L It was pleasing to see how the comments showed a strong commitment from the learners to the videoed PowerPoint visual presentation as a measure of academic skills. It would not have provided students the opportunity to express themselves particularly well, so the five minutes they might speak within a 10 minute interview became the five short reflections. As an observing teacher /researcher the students appeared to be working towards the same learning outcomes as were intended, but what was surprising were the expressions that accompanied the oral reflections.
Reference to source material: Lines 58-59: "Brockbank and McGill, 1998": not all three editors are mentioned	Lines 58-59: Beech added

Line 62: 'Huang, 2009': source not included in the 'References' Line 84: page mentioned before the authors Line 148: "Pal and Koris, 2021": In the 'References' the authors are mentioned in a different order. Line 202: "Koshy and Drane, 2021": source not included in the 'References' Lines 733-734: This source was not found as an in-text citation. The guidelines regarding the way sources of more than two authors are cited within text should be revised, because there has been some inconsistency (e.g. Swain et al. (line 62) vs Vogt, K., Tsagari, D. & Csepes, I. 2020 (line 80)).	Line 62: included Line 84: altered Line 148: changed Line 202:O'Shea et al entered inline text. First author was missing. It is in references Lines 733-734: Chen and Sato entered in rationale on pedagogically-sound assessment methods in fully online contexts; sources of more than two authors are cited Lines numbers in original text correspond to changes Line 80 Vogt, Tsagari, and Csepes 2020)

Reviewer 2	Thank you editor 2 for thorough copy editing for spaces, capitals and word division that the template forced me into accepting. I have taught word division but didn't spot the trap with automatic divisions ☹.
	I accepted and execute all the recommendations first up before addressing reviewers 1 and 3.
Line 131 132 changing 'are' to' is.'	Too far to read well in the plural although grammatically correct of course. Reordered instead . Four semesters of language study are the minimum graduation requirement for International Studies and Spanish is one of four language sequences
line 345 (sic) addressed by removing and adding a footnote	removed
Line 528 529	Sic removed
Line 359 two words run together with typo	to be my
Line 383 sic	Sic removed

Reviewer 3	Excellent editing recommendation, gracias! My apologies for the typos. It is a long story involving a very rocky island with little wifi, a small device while being away from home longer than intended due to COVID (2 extra months) and difficulties with my university

	remote desktop. I spent longer battling than writing so thanks for your patience. Thanks to your thorough questioning and comments I have made clarifications which enable this work to reach a wider audience.
Lines 22 to 32 have three purple highlights with no comment so I assume expression?	Sentence 1 Line 24 joined to sentence two by One such context… Sentence 3 reworded Line 27 Technical issues were to affect the assessment practices of most stakeholders who were using diverse platforms for the first time. For each course delivered, practitioners rationalized content, adapted assessments, and marking criteria to meet revised learning outcomes. Sentence 4 expanded to make greater links that emanated from the changed conditions due to the pandemic. By stealth, COVID directed a general overhaul of teaching towards student-centred learning. The focus on multimedia-based learning with new e-assessments that had been looming but had met resistance , was long overdue.
Line 36 which	The study reported on below…. through guided, recorded and marked oral reflection tasks, based on assessments that learners submitted during the ERT implementation period.
Line 50 Poorly-written paragraphs. Underdeveloped, both in terms of the main topic within the paragraph, as well as its relation to the broader topic.	A rewrite of the section highlighted now in purple in response to reviewer 3 Line 42 removed and ‘ for different purposes’ added to line 33 Line 47 about Brookfields Lens moved after the sentence The analysis adds to line 38-41 Line 49 and 50 about non- threating moved to methods section where student complete 5 and decide which reflection is best for me to mark . inserted and highlighted in section under ‘Course assessment tasks’
Line 47 to 62 t lacks flow and coherent transitions from one sub-topic of reflection to the next sub-topic.	Rearranged and linked

Line 69 students oriented towards 'I am not aware of a variety of English that uses the verb "orient" in this way.	I humbly claim my use is accepted in higher education and language assessment research
Line 71 to 73 active voice better	Such research provides evidence that students value the qualities they attribute towards e-assessments as they compare their experiences to an easement free context know to them from previous education in F2F non remote courses
Line 75 The main aim of the study was to uncover student orientation to e- assessment underlined no comment	I left as is as I accept the use of orientation but I am open to the final editor advice on this use of terms
Line 79 what exactly is the reference for here Language students with no previous experience of such language learning offer 78 insights on experiences of e-assessment and what they considered conducive to learning 79 (Vogt, K., Tsagari, D., & Csepes, I. 2020) which would be beneficial for course designers 80 and language instructors.	Sorry it was unclear. I have made it more explicit I hope. "….."
Inconsistent referencing style	Rectified following reviewer 2 who painstakingly detailed the minutiae to adhere to throughout and for which I am most grateful being unfamiliar with Chicago.
Line 84 86 since repeated	Removed because they were 'surprised' when ERT was implemented two-weeks into their semester.
In the handbook line 87 what handbook	LOL it is truly an anachronism from when courses were set out in a faculty handbook not online. Corrected to ' online course outline'
L 89 ''a richer and nuanced picture ''compared to what?	'A rich and nuanced picture' removed comparatives
Line 95 RQ@ seems too broad Seems to broad. What was learned in what sense? Actual content? Proportion of the ILOs? Question is vague.	As a practitioner what did I learn from adapting setting and assessing the paper based to e assessment tasks?
Line 151 Why? Explain? Were there more assessments given in this new modality than in F-2-F?	
Line 98 the teacher researcher	It is highlighted. .I can't say I so I call myself what I am . Teacher/ Researcher
Line 154 4 weeks	In words four weeks
Line 160 (n = 25)This belongs right after "Assessment practices." Assessment practices are explored here from a teacher/researcher and student perspective (n=25).	I am not reporting on 25 assessment practices. I I will hold back the info till the participants section.
Line 167 It seems to me that this is part of your "Materials and Methods."	Moved section as also recommended reviewer 1
Line 169 Above, you introduced and began using "F2F".	F2F consistent throughout after line 102
Yline 174 our appendices are titled "A" and "B" and I think you mean "A".	Changed to A & B

Lone 174 What was the purpose of this marking rubric?	The purpose of the rubric was to guide the students in the production of the reflection and for the assessor to award marks. Just as rubrics guide students on assessor expectations in other assessments, Chen and Chan (2019) experimented by incorporating a rubric as a resource for reflection. They found the rubric promoted students' level of reflection. The rubric was written in simple clear language based on issues reported on student use and understanding of reflection language as reported in the Asina et al.'s (2018) Narrative Assessment Reflection Rubric validation study in higher education.
Line 177 Why do you place this after week 4 and before week 14? Place either at the beginning of the table or at the end, as it encompasses the entire range of weeks in the table.	I have moved and placed it in the table after week 4. Which is the time the first reflection can take place.
Line 178 xplain what you mean by "knowledge transfer" and how/why you shifted away from that type of assessment.	To expand on this concept, 'knowledge transfer ' is often used in language tests in listening or comprehension where the information to respond to the question is in the text . The tasks set for the e-assessments however, were integrated tasks where responses combined different sources of information to provide a response.
Line 201 Once again, what are you referencing, here? Do these studies support discarding the use of these kinds of participants' data? Explain a little more	The study has a limitation in that it does not include the views of those who at the time of ERT did not have the technical facilities i.e. Wi-Fi, laptop, or suitable study environment during lockdown or the economic means for stability and livelihood as reported by for example, Drane et al. (2020) Mupenzi et al. (2020) or O'shea et al. (2021).
Line 196 comma	included
Line 216 ype out word when number is less than 10	Comprise / five
Line 214 The participants completed these reflections how long after they had completed the assessments? After the semester had ended? This detail is crucial, as any time that elapses after completion of the assessment, including receiving any feedback/grade, may alter the reflection or perception of the assessment and/or experience.	Rewritten response: With Ethics approval that allowed assessment tasks to be analyzed for research, Spanish 5 students were contacted via email . They could opt in for their reflections to be analyzed anonymously for research, with no further participation required of them. This data collection took place after the semester marks had been submitted and accepted by the course assessment committee, although the oral reflections had taken place during semester after assessment tasks starting in week 4.

Line 215 Why include the reflections on the oral reflections tasks Line 217 why include “partly irrelevant submissions”?	The students spoke and recorded the oral reflection. When it was transcribed it was the ‘data’ . Explanation addedPartly irrelevant submissions were included in the sample, i.e., if following the rubric students left out parts, or were completely off topic and not following the guide for a section of the reflection, the remainder could still be analysed.
Line 229 Did the students receive a mark on their oral reflection task? This practice could affect how they express themselves/what they choose to say.	The assessed reflection tasks (appendix A) and marking rubric (appendix B) offered students an opportunity to describe their perceived learning outcomes on five assessments completed during semester for a 10% of total course marks.
Line 229	Inserted clarification The students were to submit 5 audio reflections each (and per the rubric indicate their ‘ best’ one for marking)
Line 227 first person	~~I found it curious and interesting that.~~ I have removed on request of reviewer. As a practitioner carrying out research I feel comfortable with use of I. My class, My task, My rubric my research my choice? But leave it to the editor.
Line 235 line 250 appendix 1 change to A	Appendix A included
Line 251 Examples? Your reader needs to know what you deemed as not useful or irrelevant.	Off – task was coded as other’ and this is clearly set out and defined in lines 261 and 262. In coding if you include irrelevant data it skews results. I have included one example in response to the reviewer. The other two reviewers did not request it so the editor can decide if it is superfluous or helpful.. The student did not follow the guide and could not be marked agains the rubric. If the editors deem this addition unnecessary I am happy to remove it. I have followed the reviewer’s suggestion. The example below of an off-task irrelevance shows the student reflects only generally on the learning experience and the change to ERT, but does not mention the summary task completed prior to submitting the reflection. They mention that things improved for them as time went on and in fact for the remainder of the reflections the student followed the guide and rubric, so the work was included in the data as a ’reflection’.

	Translated from a transcription of spoken Spanish that has not been included as data for being off task "*The change to online classes was very difficult, in the one hand I didn't know how to use collaborate and when I did it was very hard to find a good internet connection. On the other hand, the transition to working online at RMIT at the same time had been a challenge and frequently I found myself too exhausted ager work when it was time to start class. I also felt that as if the other students were more advanced than I. It was impossible to see or read any facial expression and this did not help much either; even though this began to improve, as time went by everyone slowly kept adapting, but it wasn't as good as it could have been*."
Line 256 The table	It moved when I sent it. I have reset it to left hand margin justification
Line 257 removed ie that was repeated	The coding process was iterative, sorting sub-themes into three larger categories (see table 3) for reflective comments on doing the task; learning from the language and culture topics and reflecting on themselves, their peers and their feelings via appraisals or affective responses.
Line 272This does not sound like a logical explanation to me of *why* students would produce more comments in one category than another. The following sentence ("Also, as they record…") does sound like a more plausible explanation as to why we should not interpret frequency with importance.	I infer from my data. Discourse analysis and thematic analysis is what I do. Sorry if it is not logical. The editor is free to remove that sentence. 272, 273.
Line 279 How are you able to make that claim? You are not comparing these reflections to written reflections so you cannot make this claim.	I have analysed student Spanish advanced (C1) student ' written' reflections on course content before in detail and as I was speaking from experience, I apologise for the oversight and I inserted the reference. Hill and Ducasse 2020
Line 298 A table containing the main themes and sub-themes, either at the beginning of this section or as a summary at the end of this section, would be useful for the reader to gain at-a-glance knowledge.	that table #3 had labelled coding themes and subthemes in line 256
Line 297 rename table This is confusing	I added the missing theme 1,2 & 3 to table and reordered the reporting of the results to match exactly with a research question above each section.
Line 360 362 formatting Line 268 wrong table	Italics for translation Table 3 corrected
Line 302 What is this pronoun referring to?	The personal / interpersonal corrected

Line 306 don't understand what this sentence means.	The comments to replace pronoun
Line 310 . You should refer the reader back to Tables 2 and 4 as a reference to the types of tasks. 2. Integr/List does not appear as a type of task in Tables 2 and 4. What is this?	 Referred reader with illustrative examples, with translation and the task type (from tables 2 and 4) in brackets. Replaced with V&R/ report
Line 351 belongs at the end Line 363 . I don't see the collaboration portion of these extracts. 2. This sub-theme and its categorization are not clear.	Moved n=9 Suggested also by reviewer 1 I conflated the theme to personal and interpersonal.
Line 420 Is the lack of the accent mark a reflection of the speaker's pronunciation or an error on the author's part? Same question for all grammar errors in the transcriptions.	As it is a transcription, I provide the accents. I corrected that one and others I spotted. Grammar errors are the students. They speak that way. I did not correct their expression
Line 460 This is not so clearly categorized. Preparing ahead of time can be thought of as a learning strategy, which would place this type of extract in a different category.	
Line 571 Please see my comment regarding literature review in your Introduction section.	
Line 587 The more detailed descriptions of the tasks on which the learners reflected is crucial in the "Materials" section of the manuscript.	Completed also requested reviewer 1
Line 600 Academic papers are not a place for the researcher's feelings	
Line 627 I don't see where this evidence is provided. These comments could easily arise during F2F classes in which technology is used, either in class or at home.	

extra	
Line 26 more than 2 authors	Radha, et al. 2020).
Line 55	Mann et al. 2009).
Line 80	Vogt et al. 2020)
Line 110	(Gacs et al. 2020).
Line 200	(Drane et al. 2020; Mupenzi et al. 2020; O'shea et al. 2021
