https://doi.org/10.37546/JALTSIG.CALL.PCP2023-08

# 8

# The DECAT evaluation framework for digital authoring tools

Samuel Taylor. Kyushu Sangyo University
Ian Dagnall, Kyushu Sangyo University
David Johnson, Kyushu Sangyo University

## Abstract

This paper presents the digital educational content authoring tool (DECAT) evaluation framework, and its use in the analysis of H5P and Moodle Quiz authoring tools. The evaluation framework assessed the effectiveness and appropriateness of the two tools in the development of e-learning content provided to university students to scaffold speaking activities. It was seen as an important means of generating evidence on learning and course design by considering factors of learning, environment, and structure. The assessment of H5P and Moodle Quiz yielded a list of requirements and a rating of each requirement that suggested that H5P better met the learning requirements, while Moodle Quiz better met the environment and structure requirements and performed better overall. Consideration of how the use of the framework compared to instructors' anecdotal experience of H5P and Moodle Quiz concluded that the framework analysis largely reflected teacher experience but provided a more nuanced and fuller description of all the requirements permitted by each tool.

**Keywords:** digital authoring tools, evaluation framework, H5P, Moodle, e-learning

## Introduction

For the 21st century teacher, digital authoring tools, such as H5P and Moodle, are an increasingly viable means of creating learning content (H5P Group, 2023; Moodle, 2023). Evaluation of such tools is an important way of generating evidence to inform course design (Atwell, 2006; Raabe et al., 2015). Despite there being extensive literature on evaluation frameworks for learning materials (see Antonenko et al., 2017; Reinders & Pegrum, 2017; Salmon & Nyhan, 2013) and

e-learning systems (see Atwell, 2006; Stufflebeam, 2000; Volungevičienė et al, 2021), there is a lack of research on frameworks that combine features of both learning material and e-learning evaluations. Such an approach allows evaluation to include both learning and functionality. This research outlines the development of the digital educational content authoring tool (DECAT) evaluation framework for digital authoring tools, and shows how the framework allows teachers to evaluate a tool in relation to learning, such as input and interaction types, and non-learning aspects, such as usability, functionality, accessibility, and analytics. Description of the refinement of the framework through its use illustrates how to approach a context-specific evaluation of an authoring tool, and how this evaluation can help teachers create learning materials best suited to their needs.

**Evaluating authoring tools**

Authoring tools are software applications used to develop e-learning products. They generally use interfaces that allow for the simple manipulation and configuration of e-learning assets, reducing technical overhead by not requiring the writing of code or script in a programming editor (Berger, 2014). The use of such tools has been accelerated by the COVID-19 pandemic (Kang, 2021), which has highlighted the importance of creating detailed, clear, interactive, and user-friendly online courses (Taylor, 2022). The selection of authoring tools is important as the use of tools inappropriate for the context of use or lacking in durability can lead to the wasting of a lot of time and money (Berger, 2014). Inappropriate tools may result in ineffective instruction or insufficient support of learning, and tools lacking in durability can become incompatible with new versions of software (Berger, 2014). In satisfying the aspects of learning, support for learning, and durability, authoring tools should,

1. meet the expectations of the target audience
2. be efficient and easy to use and learn
3. be compatible with the knowledge and potential of the teacher-designer
4. include the desired interaction of resources
5. be consistent with the budget (Raabe et al., 2015).

Therefore, the evaluation of digital authoring tools is an important way of generating evidence on learning and course design and concerns the consideration of learning, environmental, and structural factors.

*Learning*

Effective learning experiences are the result of sound pedagogical design (Laurillard, 2012) and materials reflect the learning principles of the designer (Bouckaert, 2019). Therefore, any evaluation of authoring tools used to create digital learning experiences must include consideration of the learning design (Reinders & Pegrum, 2017). Following Bouckaert (2019), this research utilizes the list of pedagogic principles presented by Hadfield (2014) and proposes that those that apply to the evaluation of an authoring tool are the principles of input, output,

interaction, affective engagement, learner differences, and feedback. Input refers to the requirement for successful instructed language learning to provide extensive, rich, meaningful, and comprehensible input in both listening and reading. Output refers to the requirement to provide opportunities for both free and controlled use of the L2 for communicative, fluency-development purposes. Interaction involves input and output and is central to developing L2 proficiency and providing opportunities for communicative language use. Affective engagement refers to the requirement to present learning in such a way that learners have an attitude favorable to the language, its users, the teacher's skill in teaching the language, and their chances of success in learning the language, which will increase the likelihood of achieving communicative competence. Learner differences refers to the requirement for instruction to consider differences in learner proficiency and needs in the selection, sequencing, and presentation of material. In terms of authoring tools, this principle can be separated into issues of sequencing and adaptation. Sequencing deals with combinations of input, output interactions, whereas adaptation deals with features that change an interface based on the needs, abilities, and behavior of the user. Feedback refers to the requirement to provide feedback to learners that helps them improve their language in use (Bouckaert, 2019). Prioritizing learning in evaluation frameworks is a means of enabling a learner-centred approach to the analysis, evaluation and implementation of technology for language teaching and learning (Salmon & Nyhan, 2013).

*Environment*

Learning is supported by an environment that meets the needs of the learner (Boettcher, 2007). The role of educational technology is to afford educational tasks based on the needs of students and teachers. Technology affordances, features of an object that provide a type of interaction between the object and the agent, can indicate the usefulness of an authoring tool (Antonenko et al., 2017; Reinders & Pegrum, 2017), and the quality of the learning environment afforded by the tool. The current research utilized the categories of affordances provided by Antonenko et al. (2017) to identify spatial, temporal, navigation, and personalization affordances, as additional features of evaluating a tool that are not dealt with by consideration of learning principles. Spatial affordances include the ability to resize, move, and zoom in and out of elements within an interface. Such affordances can be grouped with other accessibility-related features, such as the provision of alternate formats for multimedia, language customization, and personalization affordances, such changing font, font size, and background color. Temporal affordances concern when, how long, and how many times learners are able to access learning content. Navigation affordances allow both linear and nonlinear movement through learning content, the usefulness of which depends on the saving of learner progress through materials. Two further important aspects of authoring tools that affect the learning environment they create are modification flexibility and the provision of analytic data. In order to provide

educators with the ability to incrementally adapt the learning environment to the needs of the learners, an authoring tool should afford the ability to edit any existing instructions, UI features such as labels, tabs, and buttons, settings after attempts have been made, content after attempts have been made, and to regrade any edited items that have been attempted. The latter three require that content is fully integrated with the provision of analytic data. Such data should meet the requirements of both learners and teachers, allowing learners to track their own progress and view attempt reports, and teachers to track learner progress, and access and download detailed learner attempt data.

*Structure*

Learning, and the support for learning is dependent on structural criteria identifiable through consideration of the organizational context. These criteria underpin the quality of the learning environment; the subject of most evaluations of e-learning systems (Attwell, 2006; Volungevičienė et al, 2021). Evaluation of the context of the learning experience is a key tenet of both learning materials and e-learning systems evaluations. For materials evaluation, context informs the development of criteria that deal with aspects of learning, environment, and structure (Tomlinson & Masuhara, 2017). For e-learning systems evaluation, context determines the needs, problems, and opportunities used for defining goals and priorities and judging the significance of outcomes (Stufflebeam, 2000). The researchers determined that the structural criteria that supported their learning environment were functionality – cross-device, cross-platform, and cross-browser – and issues related to the teacher-designer use of the technology, such as reusability, simplicity, content creation features, and cost. Functionality is often the focus of evaluations of e-learning technologies at the neglect of the learning facilitated by the functionality (Attwell, 2006). Reusability refers to the ability to edit, copy, and share content. Simplicity of interface, integration into an LMS, and content input were considered key issues, as were content creation features that afforded synchronous or asynchronous collaboration, use of an item library, batch upload of content, and real-time, post-production, and mobile view previewing of content.

**Evaluation frameworks**

The design of the DECAT evaluation framework was influenced by frameworks that were created to evaluate learning materials, inform the design and use of educational software, and evaluate e-learning environments. Reinders & Pegrum (2017) evaluate learning design with respect to technological affordances, pedagogical approaches, and affective principles. They created a 26-item framework for evaluating the learning design of mobile resources for language teaching and learning. The 26 items are split into five categories; educational affordances, general pedagogical design, L2 pedagogical design, SLA design, and affective design. Criteria are rated on a continuum from 1–5, and the framework facilitates the comparison of different resources; the resource with the higher total score

being the one with more teaching or learning potential. The design of the DECAT evaluation framework was particularly influenced by the flexibility of Reinders & Pegrum's (2017) framework. Criteria or categories which are irrelevant to a particular resource can be omitted from the analysis. Furthermore, the researchers stress the importance of the systematic evaluation of resources, including post-use evaluations to help identify obstacles to use and learner experiences with the resources.

The DECAT evaluation framework uses the three-option judgment of criteria found in the 13-item framework created by Salmon & Nyhan (2013). It was based on research into effective language teaching and learning and the role of technology in education. In evaluating software, the 13 items represent criteria by which an evaluator judges software. The range of such judgments includes, 'yes', 'no', 'unclear', and 'yes, but..' statements. Another feature of the DECAT, that of creating criteria relative to user needs, is taken from the ETAAT (Educational Technology Affordance-Ability Taxonomy) (Antonenko et al., 2017). Antonenko et al. (2017) take an affordances approach to help designers and users of educational technologies align desired user actions with the affordances of a technology. The ETAAT consists of 10 affordance types: media affordances, spatial affordances, temporal affordances, navigation affordances, emphasis affordances, synthesis affordances, metacognitive affordances, personalization affordances, adaptation affordances, and socialization affordances. Within each category, the evaluator can create criteria relative to user needs. The scale provided allows evaluators to indicate ratings of usability in relation to the target situation of use.

The Technology Enhanced Learning Accreditation Standards (TELAS) framework was developed by the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE). It is an instrument for evaluating the quality of digitally enhanced learning and teaching (DELT) at higher education institutions. The framework was developed through research and consultation with stakeholders and practitioners in the tertiary education sector and evaluates online courses across four domains: the online learning environment, learner support, learning and assessment tasks, and resources. Each domain is further divided into two standards and several specific performance criteria. Success indicators statements are used to define the meaning of each criterion and are evaluated on a scale of 'yes', 'yes, but', 'no, but' or 'no' (TELAS, 2023). The DECAT evaluation framework includes a particular strength of the TELAS framework: its use of clear and practical descriptors (Volungevičienė et al, 2021). Whereas the graphical display of results was taken from the Context Input Process Product (CIPP) model (Stufflebeam, 2000). This is also designed to assess the quality of digital learning environments and consists of four types of evaluation: context evaluation, input evaluation, process evaluation and product evaluation. It provides evaluators with easily accessible data displayed clearly in a graphic interface, allowing for the quick comparison of data sets.

## The DECAT evaluation framework

The DECAT framework was established to evaluate digital learning materials created using both H5P and Moodle Quiz to develop speaking in an EFL course at a large private university in Japan. The results of the evaluation were used to assess the effectiveness and appropriateness of the two tools in the development of e-learning content at the university. The researchers are part of a working group which is redesigning the curriculum for a listening and speaking course for 348 undergraduate non-English major students of CEFR A1–A2 English proficiency (Council of Europe, 2001). The curriculum redesign includes the creation of new e-learning materials which students complete largely outside class as part of homework provided to preview class content. Reminiscent of a flipped classroom approach, this e-learning content aims to provide students with the comprehensive scaffolding required to successfully complete speaking activities and allow for most of the class time to be used for the repetition of speaking activities that consolidate student knowledge and develop fluency and confidence.

The DECAT evaluation framework contains a total of 17 categories split into three sections: two of six categories and one of five categories. The first six categories deal with learning and instruction: L2 learning input, L2 learner output, interaction, sequencing, adaptation, and feedback. The next six categories concern supporting learning by creating a suitable environment: temporality, navigation, accessibility, modification, technical support and feedback, and analytics. The final five categories concern the structure that supports both the learning environment and learning itself: functionality, reusability, simplicity, content creation, and cost. Each category in the framework is accompanied by a question. These are provided to clarify the meaning of each category and help users think about the requirements of the authoring tool being evaluated. The evaluation sections, categories, and accompanying questions in the DECAT framework are provided in Table 1 below.

**Table 1**
The evaluation sections, categories, and accompanying questions in the DECAT evaluation framework

| Section | Category | Question |
|---|---|---|
| Learning | L2 learning input | What type of learning input is required? |
| | L2 learning output | What type of learning output is required? |
| | interaction | What type of content interactivity is required? |
| | sequencing | What sequences of input / output / interactions are required? |
| | adaptation | What kind of adaptive features are required? |
| | feedback | What type of feedback is required? |
| Environment | temporality | In what ways is control over access required? |
| | navigation | What type of navigation is required? |
| | accessibility | What type of accessibility is required? |
| | modification | What modification features are required? |
| | technical support | What type of technical support and feedback is required? |
| | feedback | What type of analytical support is required? |
| | analytics | What type of functionality is required? |
| Structure | functionality | What type of functionality is required? |
| | reusability | What type of reusability is required? |
| | simplicity | Are aspects of teacher use sufficiently simple? |
| | content creation | What type of content creation features are required? |
| | cost | Is the cost of the intended use appropriate? |

The framework is presented in an excel file and uses two sheets in the evaluation of an authoring tool, labeled 'Requirements' and 'Tool 1'. In the 'Requirements' sheet, used to specify the requirements of an authoring tool, users write any requirements related to each category. Details can be added in a section labeled 'Description'. The 'Tool 1' sheet is used to evaluate a particular authoring tool, and contains any information inputted into the 'Requirements' sheet. This allows for the evaluation of a tool in relation to the specified requirements. Three options are provided for rating a requirement: yes – the tool fully meets the requirement, yes, but insufficient – the tool partially meets the requirement, and no, not possible – the tool does not meet the requirement. The rating for each requirement is assigned a value that is used to calculate a score out of 5 for each category. A 'yes' rating is weighted as 1, a 'yes, but' rating as 0.67, and a 'no, not possible' rating as 0. Scores out of 5 are calculated so that categories containing different numbers of requirements can be compared. Using scores out of 5 for each category

means there is a score out of 30 for the learning and environment sections and a score out of 25 for the structure section, and a total score out of 85 for a particular tool. The higher the score, the better the tool performed. Scores are visualized using radar charts to aid the user in making an evaluation. Radar charts are an effective way of visualizing multivariate statistics (Chambers et al., 2018) and as such are commonly used in evaluation and quality improvement to display the performance metrics of tools such as computer programs. They are particularly effective in comparing the performance of similar applications against each other. Users of the DECAT evaluation framework can view the performance of each tool on several variables within a single radar chart, and based on this visualization of the data can decide on which tool is most appropriate for their context.

## Method

The design of the DECAT evaluation framework was investigated using two sources of data. First, the analysis of two digital authoring tools, Moodle Quiz and H5P, was undertaken using the DECAT evaluation framework. The Moodle Quiz activity is a Moodle authoring tool which allows the user to create auto-graded quizzes with various question types, including multiple choice, matching and short-answer. Moodle Quiz also allows the user to choose from a wide range of quiz settings related to feedback, time limits and takes allowed (Moodle, 2023). Moodle is an open source LMS (learning management system) which is commonly used in universities throughout Japan. It is written in the programming language PHP. H5P is an open-source authoring tool based on JavaScript. H5P is an abbreviation of HTML5 Package, and was designed to enable easy creation, sharing, and reuse of interactive HTML5 content. H5P facilitates the creation of 57 content types, which can broadly be categorized into primarily text content types, image content types, multimedia content types, and question content types. It can be integrated into an LMS via learning tools interoperability or plug-in. It is supported by commonly used LMSs, including Moodle, Canvas, Blackboard and Brightspace. (H5P Group, 2023). Introduced to Moodle in 2016, it is currently being used on over 25,000 Moodle sites (Moodle, 2023). H5P can be integrated into Moodle using two different plug-ins, the interactive content plug-in, which is represented by a black H5P icon, and the H5P plug-in, which is represented by a blue H5P icon. There are some significant differences in how content can be controlled and edited depending on which plug-in is used to upload H5P content to Moodle. The following analysis considers content uploaded using both plug-ins.

The analysis of Moodle Quiz and H5P involved creating requirements for the authoring tools and entering them into the DECAT evaluation framework 'Requirements' sheet. Each tool was then rated on each requirement by entering a value of 1 in the appropriate rating column to produce a rating out of 5 for each of the 17 categories. The second source of data was the anecdotal experiences of the researchers in using H5P and Moodle Quiz to create e-learning content. These were documented through informal discussions and used to consider the

extent to which the results of the DECAT evaluation framework analysis matched anecdotal experience. Differences between the two data were used to suggest improvements to the design of the DECAT evaluation framework.

## Results

The results of the evaluation of H5P and Moodle Quiz, displayed in Table 2 below, indicate that H5P better meets the learning requirements, while Moodle Quiz better meets the environment and structure requirements and performs better overall. The full list of requirements generated by the analysis is provided in the Appendix.

**Table 2**

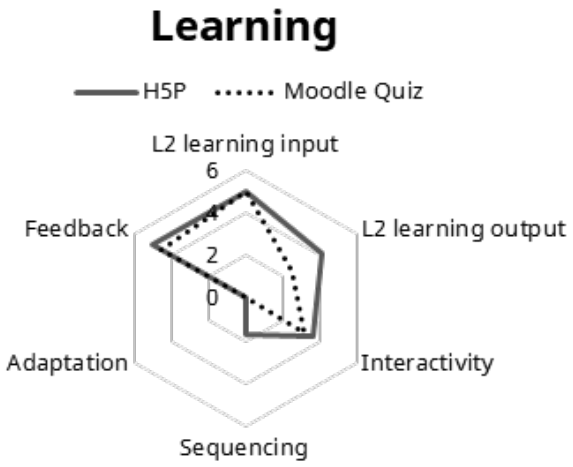
Evaluation output for H5P and Moodle Quiz

| Section | Category | H5P | Moodle Quiz |
|---|---|---|---|
| Learning | L2 learning input | 5.0 | 5.0 |
| | L2 learning output | 4.2 | 2.5 |
| | interactivity | 3.6 | 3.2 |
| | sequencing | 1.7 | 0.0 |
| | adaptation | 0.0 | 0.0 |
| | feedback | 5.0 | 4.6 |
| | Section total | 19.5 | 15.3 |
| Environment | temporality | 2.8 | 5.0 |
| | navigation | 3.3 | 5.0 |
| | accessibility | 1.7 | 1.1 |
| | modification | 3.8 | 5.0 |
| | technical support and feedback | 1.9 | 2.7 |
| | analytics | 3.9 | 4.7 |
| | Section total | 17.4 | 23.5 |
| Structure | functionality | 3.8 | 3.8 |
| | reusability | 3.3 | 5.0 |
| | simplicity | 4.5 | 3.9 |
| | content creation | 2.1 | 4.1 |
| | cost | 5.0 | 5.0 |
| | Section total | 18.7 | 21.7 |
| | Grand total | 57.14 | 62.18 |

Evaluation output for the learning categories is displayed in the radar chart below (Figure 1). Both H5P and Moodle Quiz fulfill all of the L2 learning input criteria. They perform similarly for interactivity (3.6/3.2), and feedback (5.0/4.6). H5P performs better than Moodle in L2 learning output (4.2/2.5). Both tools perform poorly in the sequencing category, H5P scores 1.7, while Moodle doesn't fulfill any of the requirements. Overall, H5P performed slightly better than H5P in the learning categories.

**Figure 1**

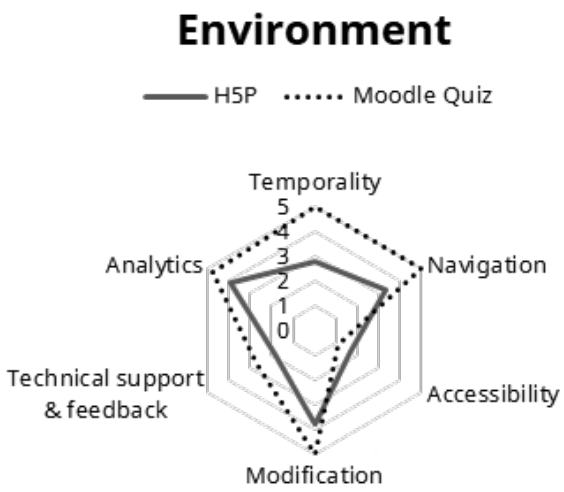
Moodle Quiz and H5P evaluation output for the learning categories




Evaluation output for the environment categories is displayed in the radar chart below (Figure 2). Overall, Moodle Quiz performed better than H5P in the environment categories. Both tools perform similarly in accessibility (3.7/3.3), while Moodle Quiz outperforms H5P in the technical support and feedback categories (2.7/1.9) and fulfills all the requirements for temporality, modification, and navigation.

**Figure 2**

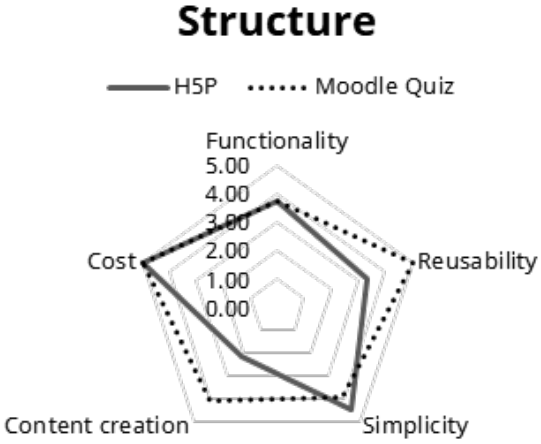
Moodle Quiz and H5P evaluation output for the environment categories




Evaluation output for the structure categories is displayed in the radar chart below (Figure 3). Overall, Moodle Quiz performs better in the structure categories. The only structure category in which H5P performs better than Moodle is simplicity (4.5/3.9). Both tools are free, so score a maximum of 5 for cost and perform the same on functionality (3.8/3.8). Moodle Quiz performs better than H5P in content creation (4.1/2.1), analytics (4.7/2.8), and reusability (5.0/3.3).

**Figure 3**

Moodle Quiz and H5P evaluation output for the structure categories



## Discussion

The analysis of H5P and Moodle Quiz using the DECAT framework showed that H5P performed better in the learning category, while Moodle Quiz outperformed H5P in the environment and structure categories, but for the framework to be useful, these results need to accurately reflect and quantify the experience of teachers using the two tools and facilitate an informed choice of the best tool for materials creation in the given context. Therefore, this discussion section will focus on the extent to which the results of the analysis using the DECAT framework reflect and improve on the anecdotal experience of the researchers in using H5P and Moodle Quiz, the extent to which the framework helps describe differences between the tools, and future modifications to the framework based on these findings.

The researchers chose H5P as a content creation tool because it was possible to create visually appealing materials with no specialist knowledge, there was a wide variety of interactive content types, and the content was responsive to mobile screen size. However, it soon became apparent that there were several limitations to creating content using H5P. While it was possible to create attractive language input activities using H5P, the lack of temporal control meant that H5P was not suitable for creating test activities, and the limited amount of data and analytics available were problematic in that it was difficult for the researchers to evaluate the extent to which activities were working as planned. Editing activities created in H5P was time consuming due to the design of the content creation interface and the limitations on copying content. In terms of navigation, H5P requires students to submit their answers upon completing each section, which can lead to problems if students have network connection problems or close the browser window before submitting their answers. In these cases students must start again from the beginning of the section.

In the researchers' experience, creating content in Moodle quiz was very different from creating content in H5P. Moodle Quiz offers little guidance on how to input content; questions and quizzes start as a blank slate and require knowledge of the syntax used in the Moodle ecosystem. However, it is simple to reuse content, and it is possible to edit and re-grade activities after they have been uploaded to the course. The teachers further appreciated access to full attempt data and analytics, which facilitates detailed analysis of how the materials are performing. However, the teachers found content created using Moodle Quiz to be visually less appealing and less responsive to mobile screen size.

The teachers' observations highlight some important differences between the tools; however, the scope of these observations is limited in that they do not consider all the requirements described in the framework, making it difficult to make an objective decision on which tool to choose. Anecdotal analysis might also attach undue importance to particular criteria. Analysis using the framework ensures that both tools are evaluated on all criteria, and that equal importance is given to all criteria. The framework analysis provides a more nuanced description

of all the requirements of each tool, and largely reflects the teachers' observations in the relevant criteria. In particular, the environment category shows the relative strength of Moodle Quiz in five of the categories, matching the researchers' observations about temporal control, navigation, analytics, and modification. However, it also shows the greater affordances offered by H5P in terms of accessibility. The more balanced analysis enabled by the use of detailed criteria is also demonstrated. Anecdotally, the teachers thought a strength of H5P was the wide variety of content types, however, the framework analysis shows that there is little practical difference between the two tools relative to the detailed criterion descriptions. A particular benefit of the framework analysis is the radar charts, which enable the user to quickly visualize the tools' relative strengths and weaknesses in each category and make the evaluation of the tools more balanced than relying on anecdotal evidence. Analysis with the DECAT framework enabled the researchers to make an informed decision on which digital authoring tool to select for content creation in their context.

While the DECAT framework has the potential to be a valuable tool in analyzing the extent to which digital authoring tools meet users' needs, there are some areas of the framework which might require modification in future versions. In the current version, 'yes, but' items are weighted as 0.67 compared to 1 for 'yes' items. In this analysis, it appears that this weighting might exaggerate the effectiveness of a tool. For example, in the learning section of the analysis, H5P appears to be relatively strong in the L2 language output and sequencing criteria, however, the difference in score is due to just one 'yes, but' rating in each criteria. Therefore, a lower weighting for 'yes, but' items, such as 0.5, might enable a more accurate analysis.

## Conclusion

Digital authoring tools are an important means of creating learning content in the modern classroom. When deciding on which tool to use, an analysis framework can help teachers make a more informed choice on which tool best meets their needs. This paper described the creation and use of a digital authoring tool evaluation framework and demonstrated how it provided a more nuanced analysis than was possible by relying solely on teachers' experience of using the tools. However, it should be noted that the criteria used in this working of the framework are specific to the context in which it was used. If the framework is used in other contexts, it is important that users create their own criteria based on their own requirements. For example, requirements for video-based instructional materials may differ greatly from those for text-based materials. Further, the framework described in this research is the first version, and there may be limitations with regards to the weighting assigned to the rating scale. Users of the current version of the framework should consider this when interpreting the analysis.

## References


Atwell, G. (2006). *Evaluating e-learning a guide to the evaluation of e-learning*. Evaluate Europe Handbook Series Volume 2.

Antonenko, P. D., Dawson, K., & Saha, S. (2017). A framework for aligning needs, abilities and affordances to inform design and practice of educational technologies. *British Journal of Educational Technology, 48*(4): 916-927. https://doi.org/10.1111/bjet.12466

Berking, Peter. (2016). *Choosing authoring tools*. https://doi.org/10.13140/2.1.4243.2002.

Boettcher, J. V. (2007). Ten core principles for designing effective learning environments: Insights from brain research and pedagogical theory. *Innovate: Journal of Online Education, 3*(3).

Bouckaert, M. (2019). Current perspectives on teachers as materials developers: Why, what, and how? *RELC Journal, 50*(3): 439–456. https://doi.org/10.1177/0033688218810549

Chambers, J. M. (2018). *Graphical methods for data analysis.* CRC Press.

Council of Europe (2020). Common European framework of reference of languages: Learning, teaching, assessment – Companion volume. Retrieved from: https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4 (accessed May 5th, 2023)

H5P Group (2023). H5P Integrations. Retrieved from: https://h5p.org/integrations (Accessed May 18th, 2023)

Hadfield, J. (2014). Materials writing principles and processes: what can we learn for teacher development? *The European Journal of Applied Linguistics and TEFL 3*(2): 71–88.

Kang, B. (2021). How the COVID-19 pandemic is reshaping the education service. In J. Lee and S. H. Han (eds.). *The future of service post-COVID-19 pandemic, volume 1, The ICT and evolution of work.* https://doi.org/10.1007/978-981-33-4126-5_2

Laurillard, D. (2012). *Teaching as a design science: Building pedagogical patterns for learning and technology.* New York, NY: Routledge.

Moodle, (2023). Moodle Usage Stats. Retrieved from: https://moodle.org/plugins/stats.php?plugin=mod_hvp (accessed May 18th, 2023)

Raabe, A. L. A., da Costa , A., & Vieira, M. F. V. (2015). Development and Evaluation of an Authoring Tool Taxonomy, in *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje, 10*(4): 204–211. https://doi.org/10.1109/RITA.2015.2486299

Reinders, H. & Pegrum, M. (2017). Supporting language learning on the move: an evaluative framework for mobile language learning resources. In B. Tomlinson (ed.). *SLA research and materials development for language learning,* 219–31. Routledge.

Salmon, J. & Nyhan, J. (2013). Augmented reality potential and hype: towards an evaluative framework in foreign language teaching. *Journal of Language Teaching and Learning, 1*: 54–68.
Stufflebeam, D. L. (2000). The CIPP model for evaluation. *Evaluation models: Viewpoints on educational and human services evaluation,* 279-317.
Taylor, S. (2022). The student experience of emergency remote teaching in compulsory English higher education classes. *Language Education and Research Center Journal, 17,* 3-31.
TELAS, (2023). *TELAS framework.* Retrieved from www.telas.edu.au/framework
Tomlinson, B., & Masuhara, H. (2017). *The complete guide to the theory and practice of materials development for language learning.* Wiley.
Volungevičienė, A., Brown, M., Greenspon, R., Gaebel, M. & Morrisroe, A. (2021). *Developing a high-performance digital education system: Institutional self-assessment instruments.* European University Association absl.

# Appendix

### The requirements generated for evaluation of the H5P and Moodle Quiz authoring tools

| Section | Learning | | | | | |
|---|---|---|---|---|---|---|
| **Category** | **L2 learning input** | **L2 learning output** | **interactivity** | **sequencing** | **adaptation** | **feedback** |
| Requirements | text | text | multiple choice | audio to speech | spaced retrieval | immediate |
| | audio | speech | drag-and-drop | audio to speech to audio | | delayed |
| | image | | cloze | | | formative |
| | video | | short answer | | | summative |
| | | | long answer | | | |
| | | | match | | | |
| | | | spell | | | |
| | | | select from list | | | |
| | | | spoken | | | |
| | | | flashcards | | | |
| | | | interactive video | | | |
| | | | true or false | | | |

| Section | Environment | | | | | |
|---|---|---|---|---|---|---|
| Category | temporality | navigation | accessibility | modification | technical support and feedback | analytics |
| Requirements | when | linear | alternate formats for multimedia | language customization for learners | technical support contacts for learners | learners can track own progress |
| | how long | nonlinear | resize, move, place elements in an interface | language customization for teachers | technical support contacts for teachers | teacher can track learner progress |
| | how many times | progress is saved | zoom | edit existing instructions | usage instructions for learners | learners can view attempt reports |
| | | | background color | edit UI features (labels / tabs / buttons) | usage instructions for teachers | teachers can view learner attempt data |
| | | | font color | edit settings after attempts have been made | comment tools for learners | teachers can access information on interpreting learner analytics |
| | | | font size | edit content after attempts have been made | comment tools for teachers | teachers can download learner attempt data |
| | | | | auto-regrading for edited items | flagging tools for learners | |
| | | | | | flagging tools for teachers | |

| Section | Structure | | | | |
|---|---|---|---|---|---|
| Category | Functionality | Reusability | Simplicity | Content creation | Cost |
| Requirements | cross-device | editability | interface simplicity | synchronous collaboration | within budget |
| | cross-platform | copyability | integration simplicity | asynchronous collaboration | |
| | cross-browser | shareability | learning content input simplicity | item library | |
| | error reports | | | batch upload | |
| | | | | real-time preview | |
| | | | | post-production preview | |
| | | | | mobile view preview | |

# Author bios

**Samuel Taylor** received his MA in TESOL from the Institute of Education, University College London in 2017. His research interests include speaking, student perceptions of education, and educational technology.

**Ian Dagnall** obtained an MA in TESOL from Lancaster University in 2020. His teaching and research interests include learner engagement, materials development, and language testing.

**David Johnson** has teaching and research interests including extensive reading, curriculum design and educational technology. Originally from Sheffield in the UK, he obtained his MA from Sheffield Hallam University in 2014, writing his thesis on learner demotivation.