IJSRM
INTERNATIONAL JOURNAL OF SCIENCE AND RESEARCH METHODOLOGY
An Official Publication of Human Journals
HUMAN

Human Journals
Research Article
December 2020 Vol.:17, Issue:2

Three-Dimensional Model for Quality Assessment of Digital Learning Resources

IJSRM
INTERNATIONAL JOURNAL OF SCIENCE AND RESEARCH METHODOLOGY
An Official Publication of Human Journals
HUMAN

Ozonias de Oliveira Brito Junior*[1], Yuska Paola Costa Aguiar[2], Hermano Perrelli de Moura[3]

[1 3] Federal University of Pernambuco, Center of Informatics, Brazil [2] Federal University of Paraiba, Department of Exact Sciences, Brazil

Submitted: 07 November 2020
Revised: 27 November 2020
Accepted: 17 December 2020

HUMAN JOURNALS

www.ijsrm.humanjournals.com

Keywords: Digital Learning Resources (DLRs); Software Quality Models; Quality of Use Models; Interdisciplinarity in Human-Computer Interaction

ABSTRACT

The Digital Learning Resources (DLRs) are mediators in the teaching and learning process. Given the importance of ensuring the quality of such resources, the proposition of approaches for their assessment is abundant in national and international literature. Also, considering the plurality of aspects involved to define the quality of DLRs (technical aspects, usability and user experience, accessibility, learning, teaching, etc.), the existing propositions are heterogeneous and are not supported in a macro and uniform structure. On the one hand, if the absence of homogeneity implies the wealth of perceptions about the subject, on the other hand, it impacts the understanding about what is the quality of a DLR, or about what criteria can be adopted to guide this analysis. To maximize the amplitude of the existing propositions regarding the complementarity of their criteria, as well as to minimize the lack of adoption of a common language and a meta-organization according to quality models, this article presents the result of the analysis, categorization, and compilation of a set of 26 approaches for the assessment of DLRs, from the application of the Design Science Research methodology. As a result, there is a set of 49 criteria for the assessment of the quality of DLRs, organized in categories of the dimensions of Software Quality, Quality of Use, Pedagogical Quality, and there are also hybrid criteria. This overview allows identifying what criteria are relevant to carry out a three-dimensional assessment of the quality of DLRs.

INTRODUCTION:

In the most recent years, the migration from traditional teaching methods to alternatives that include technology such as auxiliary and mediating tools has been notorious. With the great technological advance and the popularization of the internet, Digital Learning Resources (RDAs) provide the opportunity for a new form of teaching and learning [1]. For [2], these resources are increasingly being used to support the teaching and learning process, playing a relevant role in motivating and engaging the student more, both inside and outside the classroom.

The RDA is a product designed to serve as a means to implement pedagogies based on the use of the computer and contribute to achieving pedagogical objectives [3]. According to [4] [5] [6] [8], RDA is a digital device or resource that was planned and developed for didactic-pedagogical purposes, representing a content mediator and facilitator of the teaching and learning processes, bringing the practical development of school knowledge in a playful and pleasurable way.

According to [9] [10] [11] [12] [13] [14], the use of RDAs can: provide the student's ludic learning; improve the interpretation of theoretical content from its practical application; stimulate reasoning and critical thinking; allow the possibility of adaptation to the students' different learning rhythms; enable students to learn from their mistakes; reduce the passivity of learners; improve psychomotor skills; provide motivation to the student; develop habits of persistence in overcoming challenges and developing tasks; and, consequently, enrich the pedagogical practice with the use of multimedia resources.

However, the adoption of an RDA requires care, as several factors can interfere with the objectives intended for its use, for example, insertion of the RDA in a wrong educational context [15], inconsistency between the addressed and the available content via the RDA [15], usability problems [16], poor structuring and/or representation of didactic content [17], lack of pedagogical and didactic grounds adequate to the profile of the learners, and functioning problems (bugs) [18].

Therefore, it is worth noting that to achieve the aforementioned benefits in the application of RDAs, these (i) must be pedagogically adequate, considering the objectives established by the teachers; (ii) have quality according to the precepts of Software Engineering and (iii) that your project has been directed to the characteristics, needs, and skills of its end users, as recommended by Usability Engineering.

The incidence of inadequacies, in any of these dimensions, can compromise the use of the educational resource and, consequently, its main purpose of enhancing the teaching and learning process. Therefore, the adoption of these resources must be preceded by an evaluation, so that their quality is verified considering the three dimensions involved in the use of RDA in the educational context of teaching and learning - considering also its adequacy about the context in which it intends to apply it [18].

In this perspective, national and international literature presents a range of approaches (methods, techniques, methodologies, instruments, tools, others) that can be used in the evaluation of the quality of RDAs. This scenario highlights the interest of the scientific community in the development and evaluation of RDAs, given the trend of using these increasingly frequent resources. The attention directed to these resources aims to facilitate the teaching and learning process in favor of the most diverse areas of knowledge and levels of education.

The diversity of available approaches is a consequence of the wealth of perceptions about the subject. However, on the other hand, the heterogeneity of propositions can impact the understanding of what is meant by RDA quality, or even, on what criteria it can be adapted to guide an assessment of this nature. The existing proposals are heterogeneous, concerning the dimensions evaluated: technical aspects of the software, usability and user experience, accessibility, learning, teaching, etc., as well as about the number of criteria considered to compose the dimension evaluation above. Some approaches have 3 criteria, while for others, this amount is 20 criteria.

Another characteristic that deserves to be highlighted is that the approaches are not always based on a macro and uniform structure for their presentation. In other words, although there are established models in the literature about software quality, quality of use, and pedagogical quality, these have not been considered as guidelines for existing approaches. As a consequence, a whole terminology, and the associated concepts, which are commonly used in the context of Software Engineering, Human-Computer Interaction, and Pedagogical, has not been reused.

However, the diversity of approaches presents several problems related to the lack of standardization, such as the existence of different techniques for evaluating RDAs; the definition by the approaches of a specific scope for the evaluation of criteria - Software Quality (SQ), Quality of Use (UQ) and Pedagogical Quality PQ); and the use of different criteria (technical aspects, usability and user experience, accessibility, learning, teaching, didactics, etc.), which are classified according to categories defined by their authors (hybrid criteria).

This diversity indicates the relevance of the theme and leads to reflection on the lack of consensus for the evaluation of RDAs [5] [7] [11] [17] [19]. Also, this plurality can be an additional barrier in the task of evaluating RDAs, since (i) the evaluator is unaware of the totality of existing approaches, being necessary (ii) to identify them and carry out a process of (iii) selection to maximize the outcome of the assessment. With that being done, the evaluation process can be costly, as the approaches do not (iv) have equivalent rating scales, making it impossible for a mix of approaches to generate (v) a final evaluation score. In another scenario, if two RDAs, with simulated pedagogical purposes, are evaluated through different approaches, the (vi) results obtained are not necessarily subject to comparison - because of the great heterogeneity mentioned. Still, considering the dimensions involved for this type of resource (software quality, quality of use, and pedagogical quality), different profiles of evaluators are required because of the need for complementary perspectives. Therefore, the lack of information on the prior knowledge needed by the RDA evaluators stands out as problematic (vii).

The aforementioned problems have negative consequences for the RDA evaluation process, such as the evaluation of RDAs by an inadequate approach; generation of distinct results that compromise the quality of the software; use of a particular approach for a different appraiser profile for which it is recommended, among others.

Therefore, bearing in mind the importance of enhancing the effectiveness of existing RDA assessment approaches, considering their plurality in their uniqueness; as well as minimizing the inconveniences arising from this heterogeneous diversity, the objective of this research was to identify, select and analyze a set of 26 approaches for evaluating RDAs to compile the existing criteria and map them according to software quality models [20] [21], of use [22] [23] [24] [25] and pedagogical [26] [27], recommended in the literature. As a result, there is a proposal for a three-dimensional model for evaluating RDAs validated by 32 Software Engineers, 45 usability specialists, and 4 educators. The model consists of a total of 49 criteria, adopting a common language and a meta-organization, which also includes the description of the competencies required for the profiles of evaluators.

Therefore, it is believed that the beginning of standardization for the evaluation of RDAs consists of the analysis of existing approaches to identify their characteristics and propose a common language and meta-organization. Motivated by this, this work presents the analysis of a set of approaches for the evaluation of RDAs found in national and international literature and proposes a compilation of the 26 approaches as a model for three-dimensional evaluation. Also, the work presents the validation of the model proposed by specialists in software quality, quality of use, and pedagogical quality.

MATERIALS AND METHODS:

Methodologically, this research work is classified, as far as nature is concerned, as applied research, as it aims to generate new and useful knowledge for the solution of real problems [23]. As for the research logic, it is classified as inductive, because it is based on inductive arguments, in which the researcher infers theoretical concepts and patterns, from observed data, developing general conclusions about specific observations, moving from specific to the general [23]. As for the objectives, it is classified as exploratory research, as it seeks new principles to improve current approaches [28]. As for the research process, it uses a mixed approach [28] [29], combining qualitative and quantitative research methods. *Design Science Research* (DSR) [30] was utilized as a research method. According to [30], DSR is a method oriented to solving problems with the construction of a solution through the application of an artifact (method, model, guideline, resource, software).

The DSR was conducted in 4 (four) phases: **PHASE 1:** Identification and Analysis of Approaches to Assess RDAs in order to know which dimensions and criteria are proposed by them, conducted from a literature review (snowball / automatic search); **PHASE 2:** Compilation of existing criteria in the identified approaches in order to eliminate duplicates and to identify criteria with the same objective and different nomenclatures; **PHASE 3:** Mapping of the criteria compiled according to the Software Quality [20] [21], Quality of Use [22] [23] [24] [25] and Pedagogical Quality [26] [27] models, recommended in the literature in order to propose a three-dimensional model for evaluating RDAs, composed of the complementarity of the approaches considered in the corpus, organized in the dimensions mentioned above; and **PHASE 4:** Validation of the model with specialists in the dimensions of quality to analyze its consistency, integrity and ambiguity, through the application of online questionnaires and semi-structured interviews.

PHASE 1: Identification and Analysis of Approaches to Assessing RDAs (Literature Review). The literature review was conducted in three stages. Step 1, manual search in the context of national publications, which resulted in the identification of a set of 14 approaches for evaluating RDAs from 63 articles. Details of this stage and the results achieved are available in [11]. Step 2, carrying out the snowball process [31] for the corpus initially analyzed in Step 1, resulting in a quantity of 8 new articles for analysis, of which 4 approaches for the evaluation of RDAs were added to the corpus of this study. Step 3, Systematic Literature Review (RSL) [28] to increase the body of knowledge on the topic, including research, carried out and published in international media. The execution of the RSL was performed through a pre-established protocol (Figure 1).

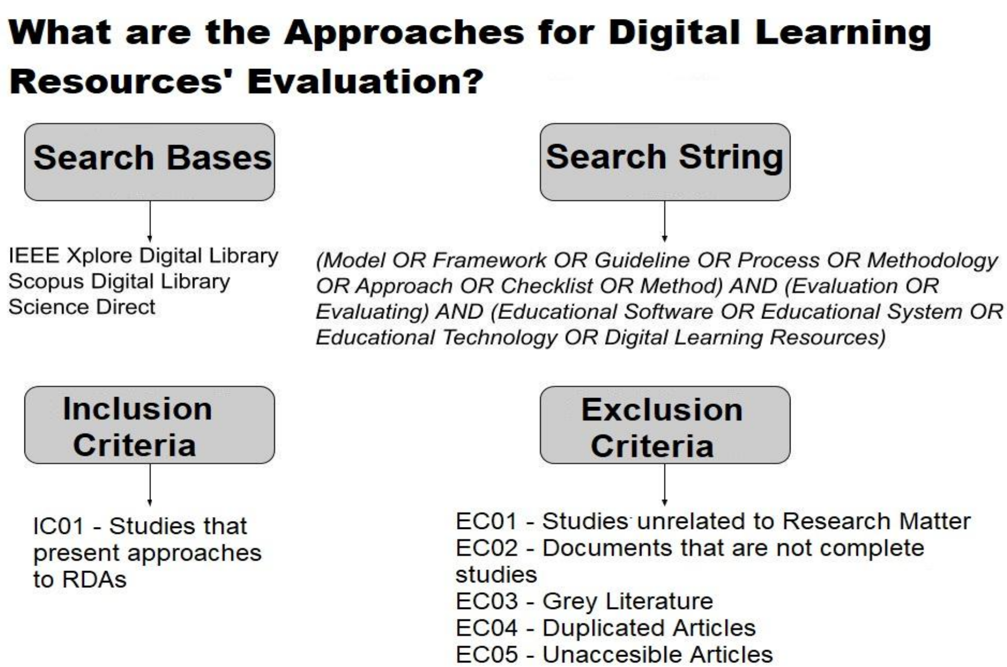


Figure No. 1: Systematic Literature Review Protocol

Source: Prepared by the author

To conduct RSL, the search string was performed on the search bases (IEEE, Scopus, and Science Direct) by a Ph.D. student and two professors and specialist researchers in Usability, Human-Computer Interaction, Software Engineering and Evaluation of RDAs. As each base has particularities regarding the execution of Strings, these needed to be calibrated.

As a result, 102 articles were returned (IEEE: 27; Scopus: 12 and Science Direct: 63 articles). After applying the inclusion and exclusion criteria, after reading the titles and abstracts of all articles, 18 articles remained for complete analysis (IEEE: 4; Scopus: 5 and Science Direct: 9). These articles were read in full, with reapplication of the inclusion and exclusion criteria. At the end of this selection, 9 articles were integrated into the corpus of this study (IEEE: 2; Scopus: 2 and Science Direct: 5).

It is worth mentioning that the final corpus composed of 26 approaches for the evaluation of RDAs is the result of the three review steps mentioned above. Therefore, there is an expressive total of 172 national and international articles consulted to obtain the corpus, namely: 62 resulting from step 1 (manual review, national context), 8 resulting from step 2 (snowball) and 102 resulting from step 3 (SLR).

PHASE 2: Compilation of Criteria proposed by the 26 Approaches Identified in the Literature Review. For the analysis of the 26 approaches identified in the Literature Review, the respective original publications were retrieved to know the proposition from the source. With access to this document, the authors, the place and year of publication were filtered, to understand the relevance and topicality of the approaches.

Once the approaches were selected, their characterization was carried out considering the elements: (i) nature of the evaluation: (ii) quality aspects evaluated by their criteria (software, use and pedagogical); (iii) instruments for data collection; (iv) number of criteria for each approach according to each dimension highlighted by the author of the article; (v) types of evaluation scales used; and (vi) results (diagnosis) obtained by each approach.

In addition to this characterization, information was also extracted to eliminate duplicate items; and to identify distinct nomenclatures for the same concept. We also filter the definitions of each criterion that formed the approaches to allow a better understanding of what the objective of that criterion would be in the evaluation. Finally, we surveyed this information and started the next phase.

PHASE 3: Mapping of Criteria Compiled in Quality Models. Phase 3 aimed to map between the 26 approaches selected in Phase 1 with the Software, Use and Pedagogical quality models, in addition to the hybrid criteria. Initially, for each model, the existence of repeated criteria was identified. For SQ, the 27 criteria available in ISO 9126-1 [20] were compared with 11 in the

Swebok Guide [21], of which 16 were equivalent, leaving 21 Software Quality criteria. For UQ, the 7 criteria available in ISO 9241-1 [22] were compared with the 10 usability heuristics defined by Nielsen [23], the 8 proposed by Shneiderman and Plaisant [24] and the 8 suggested by Bastien and Scapin [25]. The Quality of Use model, with no duplication, now comprises 12 criteria.

For the pedagogical precepts of quality, the proposals of [26] with 19 criteria and of [27] with 14 criteria were considered, contemplating 27 criteria after removing duplicates. Duplicate criteria were allocated to only one of the categories (most relevant). In addition to this survey, there was also an intersection between the three quality models to identify whether there is any correlation between the criteria presented in each of them.

With the characterization of the 26 completed approaches and the selection of terms (criteria) finalized for each quality model, a detailed mapping was performed, considering the quantitative ones related to the highest and lowest number of criteria for each approach. The mapping analyzed 60 criteria from the three quality models with 724 criteria from all 26 approaches.

PHASE 4: Validation of the model with specialists in Quality Dimensions. Phase 4 aimed to validate the proposed model through the opinion of experts in each of the three dimensions of quality addressed. To validate the model, online questionnaires and semi-structured interviews were used. The dimensions of Software Quality, Quality of Use and Hybrid Quality were validated through questionnaires, which were structured in 4 sections, namely: (i) consent form for research participation; (ii) identification of the respondent's profile (training and professional performance of the respondents); (iii) assessment of the adequacy of the model with the presentation of the criterion/description; and (iv) thanks for participation. For each model criterion, the respondent should indicate agreement or disagreement, being possible for the second case, to give suggestions for changes and/or improvements. The HQ dimension criteria were assessed in the SQ and UQ questionnaires according to their arrangement in the respective dimensions.

The dimension of Pedagogical Quality was assessed through semi-structured interviews with the participation of education specialists. This strategy was adopted considering the wealth of information that can be discussed and collected - essential for understanding concepts and respective descriptions - as a way of balancing the fact that the model's proposal was conducted by experts who do not master the pedagogical dimension. Three virtual 90-minute meetings were held (06/20, 07/04, 07/06 2020), with the participation of 4 educators (1 specialist, 1 doctoral student, 2 doctors). The objectives of the meetings were: (i) analysis of category nomenclatures and their arrangement in the proposed model, (ii) analysis of criteria nomenclatures and their allocation within each category and (iii) analysis of the description of each criterion (in two moments).

RESULTS AND DISCUSSION:

The evaluation of RDAs is based on approaches that contemplate different characteristics; these, in turn, focus on different criteria for evaluation. The national and international academic literature presents a variety of approaches for the evaluation of RDAs, among which the present work selected a set of 26 approaches, the description of which follows presenting each one of them:

1. Reeves method [32] considers the evaluation of the RDAs from 24 criteria, from the pedagogical perspective (4) and user interface (14). This method is suitable for evaluating completed RDAs and for the specific profile (evaluators) of teachers, but also developers.

2. MAQES (Methodology for Assessing the Quality of Educational Software) [33] adopts a checklist, with 9 technical criteria, related to the usability and conceptual reliability of RDAs, where it guides software developers and users (teachers and students) with a manual for the evaluation of RDAs.

3. Mucchielli's technique [34] uses a checklist with 10 Software Quality criteria to be verified during its evaluation as an evaluation tool; it is recommended to guide usability experts and developers to evaluate RDAs.

4. LORI (Learning Object Review Instrument) [35] is normally used for the evaluation of Learning Objects (LOs) based on 9 criteria regarding the quality of LOs. This instrument is indicated for the profile of SQ specialists and developers.

5. ESEIT (Educational Software Ergonomic Inspection Technique) [36] adopts several criteria with a particular focus on ergonomics, guiding the evaluator to carry out an inspection in the RDA. This technique is recommended for conducting mixed assessments, and used by ergonomics experts and teachers.

6. Rocha Method [37] is used for the evaluation of RDAs of the types of Learning Object and Educational Sofware based on 23 criteria. It is indicated to carry out evaluations by usability specialists and software developers.

7. MAQCES (Methodology for Assessing the Quality of Children's Educational Software) [38] adopts several criteria (14) related to the SQ and UQ for assessing children's RDA, it is recommended for use by teachers and usability specialists.

8. QAIEMS (Quality Assessment Instrument for Educational Mathematics Software) [39] is used for the evaluation of RDA based on 55 criteria referring to SQ, UQ and PQ. This instrument is recommended for use by teachers, programmers and usability specialists.

9. EGEM (Educational Games Evaluation Model) [40] considers the evaluation of RDA from 43 criteria from the perspective of UQ and PQ, it is recommended to be used by students and teachers.

10. ESE Method (Educational Software Evaluation Method) [41] adopts several criteria (34) related to SQ, UQ and PQ for RDA evaluation, it is indicated to be used by students and teachers.

11. MAEP (Interactive Ergopedagogical Evaluation Method for Computerized Educational Products) [42] is used for the evaluation of computerized educational products based on 36 criteria related to quality. This instrument is indicated for the profile (evaluators) of usability specialists, programmers, students and teachers.

12. IAJEEEF (Instrument for Evaluating Educational Electronic Games in Elementary School I) [43] is used for the evaluation of RDAs (educational games) based on 48 criteria related to the quality of educational games in the educational context. This instrument is recommended for evaluations by teachers and students.

13. STESA (Specialist Tool for Educational Software Assessment) [44] considers the RDA assessment based on 24 criteria referring to SQ, UQ and PQ. This tool is suitable for the profile (evaluators) of students, teachers, usability specialists and programmers.

14. ETEF (Educational Tool Evaluation Form) [45] adopts 37 criteria to assess the dimensions of Quality of Use (11 criteria) and Pedagogical Quality (26 criteria). This form is suitable for the profile (evaluators) of usability specialists and teachers.

15. MEEGA + (Model for the Evaluation of Educational Games) [46] is used to assess the quality of RDAs for computer education based on 31 criteria related to UQ and PQ (focus: to evaluate the perception of quality in terms of player experience and perception of learning). This model is recommended for use by students and teachers.

16. EMDER (Evaluation Model of Digital Educational Resources) [47] considers the evaluation of RDAs from 20 criteria from a pedagogical (academic, pedagogical and didactic) perspective and technical (design, navigation and usability), considering PQ and UQ criteria. This model is suitable for teachers and usability specialists.

17. ESHTRI Model (Educational Software Hierarchy Triangle Model) [48] adopts a hierarchical triangle model with 13 criteria for evaluating the SQ of RDAs. This model is recommended for use by programmers only.

18. TUP Model (Evaluation Model Technology, Usability and Pedagogy) [49] is used for the evaluation of RDAs based on 15 criteria referring to SQ, UQ and PQ. This model is suitable for the profile (evaluators) of usability specialists, programmers, students and teachers.

19. MARDA (Method for Evaluating Digital Learning Resources) [50] adopts a questionnaire, with 20 UQ criteria for the evaluation of RDAs. This model is designed to guide usability experts in the process of evaluating categories related to text, color, navigation and layout.

20. GUEDES (Guideline for the Evaluation of the Design of Educational Software) [51] is used for the evaluation of RDAs based on 14 criteria related to Quality of Use, specifically focusing on the evaluation of accessibility with regard to the use of users. This guideline is recommended for evaluations by usability experts and teachers.

21. CASE (Checklists for the Evaluation of Educational Software) [52] considers the evaluation of RDAs from 15 criteria from the perspective of UQ and PQ, it is recommended to be used by usability specialists, students and teachers.

22. USEES (Usability Scale for Evaluating Educational Software) [53] is used for the evaluation of RDAs based on 37 criteria related to SQ, UQ and PQ. This scale is indicated to guide usability specialists, programmers and teachers.

23. PETESE (A pedagogical ergonomic tool for educational software evaluation) [8] adopts 68 criteria related to SQ, UQ and PQ, which is indicated for usability specialists, teachers and programmers.

24. UCMIA (Usability Checklist for Mixed Interface Assessment) [54] considers the assessment of RDAs based on 40 criteria from the perspective of UQ, it is recommended to be used by usability experts.

25. ESES (Educational Software Evaluation Scale) [55] adopts 47 criteria related to the SQ, UQ and QP. This scale is recommended for use by usability experts, teachers and programmers.

26. CAPES (Checklist for Assessing the Potential of Educational Software) [56] is used for the evaluation of RDAs based on 21 criteria referring to SQ, UQ and PQ. This checklist is intended to guide usability experts, programmers and teachers.

The 26 approaches presented have specific particularities, according to the dimension of evaluation and the types and quantities of criteria that are addressed in each one. For example, the approaches of Mucchielli, MAJE, IAJEEEF, ETEF, MEEGA +, EMDER, MARDA and GUADESE do not contemplate any criteria of the Software Quality model. The ESHTRI model and MAQSE approaches contemplate 10 and 8 criteria, respectively, of this model.

The MAQSE and LORI approaches do not include any criteria of the Quality of Use model, while the MAEP and PETESE approaches include 8 criteria of this model. The ROCHA, ESHTRI Model and GUADESE approaches do not include any criteria of Pedagogical quality, while the IAQSEM and MEEGA + approaches include, respectively, 12 and 10 criteria of this model.

It can be seen that some approaches have a wide range of criteria, covering different dimensions in the evaluation of RDAs; therefore, other approaches assess a limited number of criteria. Therefore, the configurations highlighted by the approaches instigate reflection on the need to work on the complementarity of the approaches in a compilation of them.

From the corpus (26 approaches) presented, mapping was carried out with the quality models established in the literature, and the analysis resulted in the presentation of a model formed by three dimensions: Software Quality (Figure 2), Quality of Use (Figure 3) and Pedagogical

Quality (Figure 4), in addition to a Hybrid Quality dimension (Figure 5), whose criteria involve more than one dimension.

The developed three-dimensional model aims to identify whether the RDAs are following the standards and specifications of Software Engineering so that they do not present problems related to their performance. This assessment consists of identifying compliance with the RDA's functional and performance requirements and analyzing the implicit characteristics that are expected during its execution.

The model also seeks to identify the quality of use by analyzing the ability of the software product to be used by specific users to achieve specific objectives, without compromising their experience and use. As the RDAs also focus on pedagogical quality; the proposed model seeks to evaluate aspects related to learning, teaching, didactics and pedagogical content established in school curricula and educational institutions.

The dimensions highlighted in the mapping are separated into categories; each category consists of a set of criteria - these coming from the 26 approaches to assessing RDAs, previously analyzed and compiled. For each criterion mapped in the model, the frequency of its occurrence was recorded as a proposition in the 26 analyzed approaches. Although this association does not indicate that the most frequent criteria are more important than the less frequent ones, it can be said that there is a greater occurrence in the literature of the need to evaluate these criteria in the RDAs.

In total, the model for three-dimensional analysis of the quality of RDAs is composed of 49 criteria, organized in three dimensions that total 17 (seventeen) categories. This model is useful to define which criteria should be considered to assess the quality of RDAs considering their particular characteristic of three-dimensionality. We sought to adopt a meta-organization according to the language commonly adopted by Software Engineering, Usability Engineering and Pedagogical Foundations.

The Software Quality dimension (Figure 2) consists of 3 (three) categories and 17 (seventeen) criteria. The three categories and the number of criteria that compose it are Software Evolution (7), Software Design (6), System Security (4). Among the dimensions, we have that the Software Evolution category is the most numerous with 7 criteria and System Security the least with only 4. As for the occurrence of the criteria of this dimension in the corpus analysis, considering the

26 approaches, we have the criterion Fault Tolerance mapped in any approach and the Adequacy, Resources and Apprehensibility criteria are mapped in seven of them. Related to the characteristics of this dimension, the profiles indicated to carry out its evaluation are software developer, programmer, software tester and specialists in software engineering.

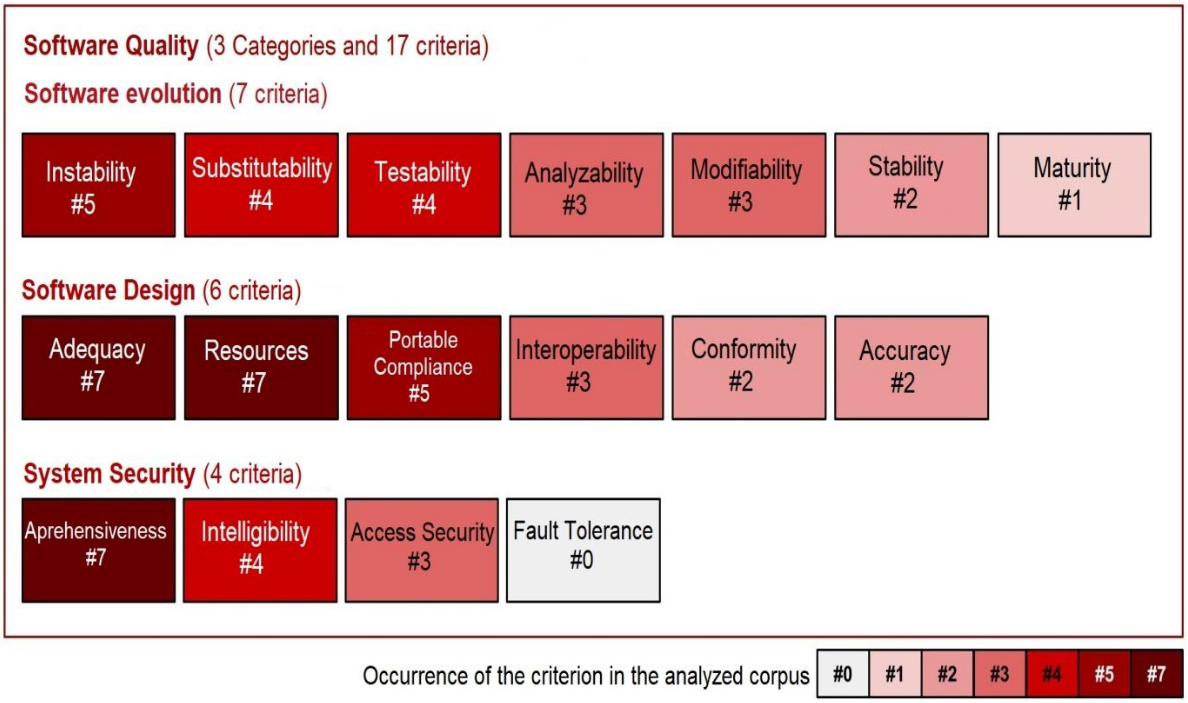


Figure No. 2: Software Quality Dimension (Source: Prepared by the author)

The Quality of Use dimension (Figure 3) is made up of 4 (four) categories and 10 (ten) criteria. The four categories and the number of criteria that compose it are User interface (4), User interaction (2), User experience (2), and Error management (2). Among the dimensions, we have that the User Interface category is the most numerous with 4 criteria and the other 3 the least, with only 2 each.

Regarding the occurrence of the criteria of this dimension in the analyzed corpus covering the 26 approaches, we have that the Criteria for error prevention and support to recognize, diagnose and recover errors was mapped in four approaches and the criteria Flexibility and efficiency of use and aesthetic and minimalist design in 21 and 16 approaches, respectively, standing out as the most cited criteria in the approaches.

This number of occurrences of these criteria in the approaches highlights that they are essential in the analysis of the usability of RDAs. Related to the characteristics of this dimension, the profiles indicated to carry out their evaluation are students, teachers, specialists inaccessibility, specialists in User experience, specialists in heuristic evaluations, specialists in human-computer interaction, and specialists incommunicability.

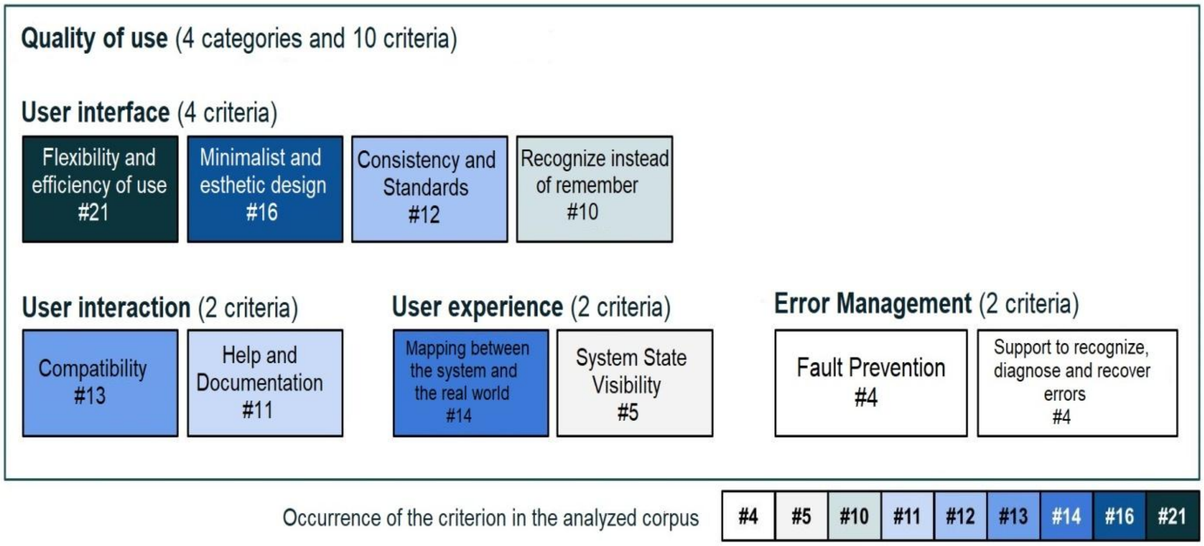


Figure No. 3: Quality of Use Dimension (Source: Prepared by the author)

The Pedagogical Quality dimension (Figure 4) consists of 6 (six) categories and 15 (fifteen) criteria. The six categories and the number of criteria that compose it are Relevance of the content (3), Motivational and responsive resources (3), Pedagogical content (3), Fundamentals and pedagogical objectives (2), Previous knowledge (2), and Epistemological clarity (2). Among the dimensions, we have 3 categories with 2 criteria; each of them and 3 categories with 3 criteria, this balance in the number of criteria by categories highlights the importance of evaluating all the pedagogical aspects of an RDA.

As for the occurrence of the criteria of this dimension in the analyzed corpus covering the 26 approaches, we have that the criteria Record of student performance and an indication of previous knowledge were mapped in only 3 approaches and Adequacy of the RDA to the target audience and the school curriculum, Adequacy of the RDA to the content worked on and Feedback that is encouraging and free of negative charge through inadequate responses were mapped into 16, 15 and 10 approaches, respectively. Related to the characteristics of this dimension, the profiles indicated to carry out its evaluation are: specialists in educational technologies, students, teachers, pedagogues, educational supervisors, and specialists in the area of education in general.

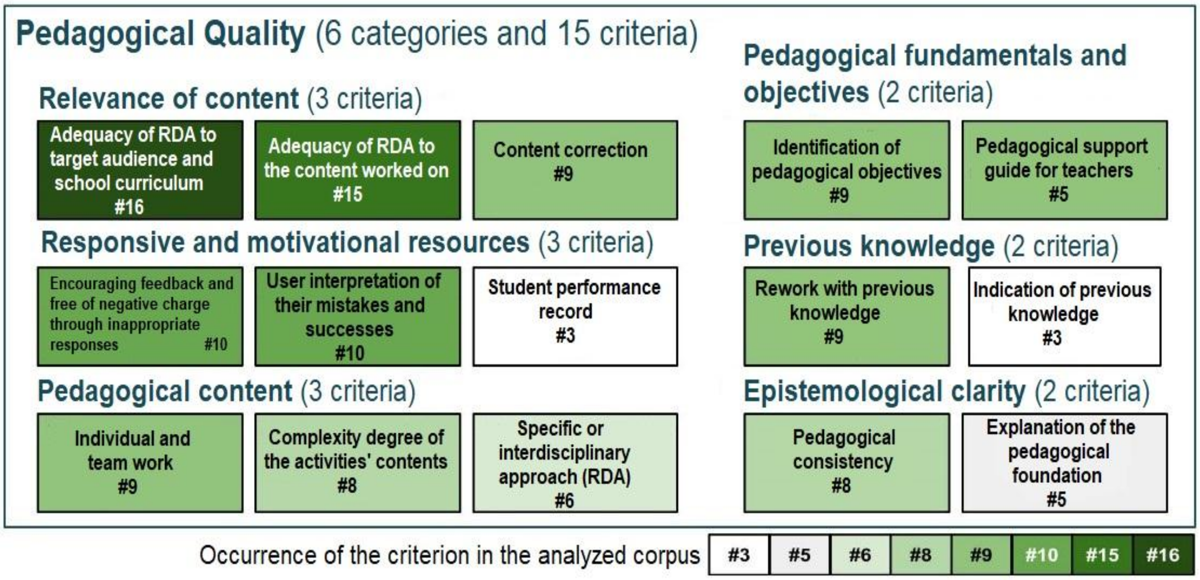


Figure No. 4: Pedagogical Quality Dimension (Source: Prepared by the author)

The Hybrid Quality dimension (Figure 5) is made up of 4 (four) categories and 7 (seven) criteria. The four categories and the number of criteria that comprise them are Software Design and User Experience - SQ + UQ (1 criterion), Software Evolution and User Experience - SQ + UQ (1 criterion), System Security and Experience User - SQ + UQ (2 criteria) and User Interaction and Motivational and Responsive Resources - UQ + PQ (3 criteria). Among the dimensions, we have that the category User Interaction and Motivational and Responsive Resources is the most numerous with 3 criteria, and the categories Software Design and User Experience and Evolution of Software and User Experience are the least numerous with 1 criterion each.

Regarding the occurrence of the criteria of these dimensions in the analyzed corpus covering the 26 approaches, we have that the criterion Support for the user to recognize, diagnose and recover errors was mapped in only 4 approaches and the criteria Freedom and User Control, Attractiveness and Pedagogical Challenges were mapped in 18, 15 and 11 approaches, respectively. Related to the characteristics of this dimension, the profiles indicated to carry out its evaluation are those that correspond to the dimensions involved in the joining of the categories.

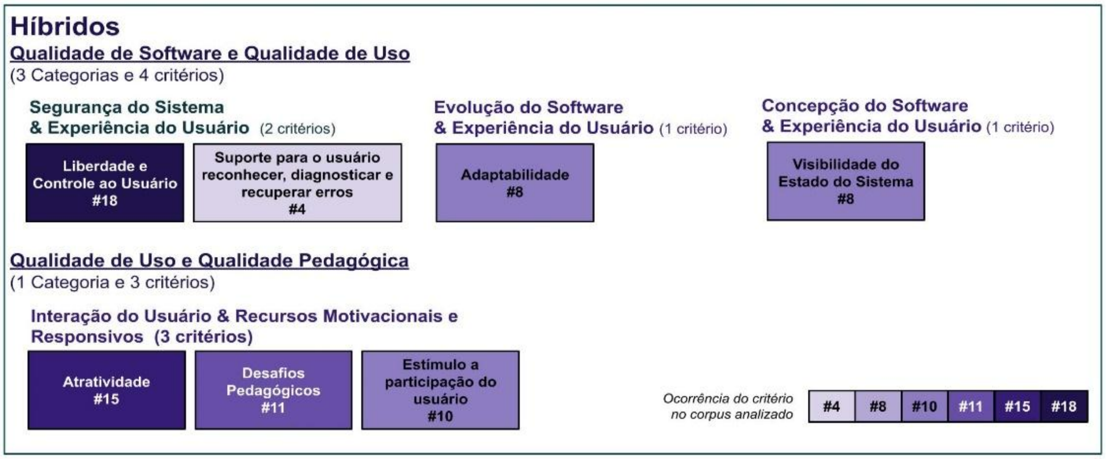
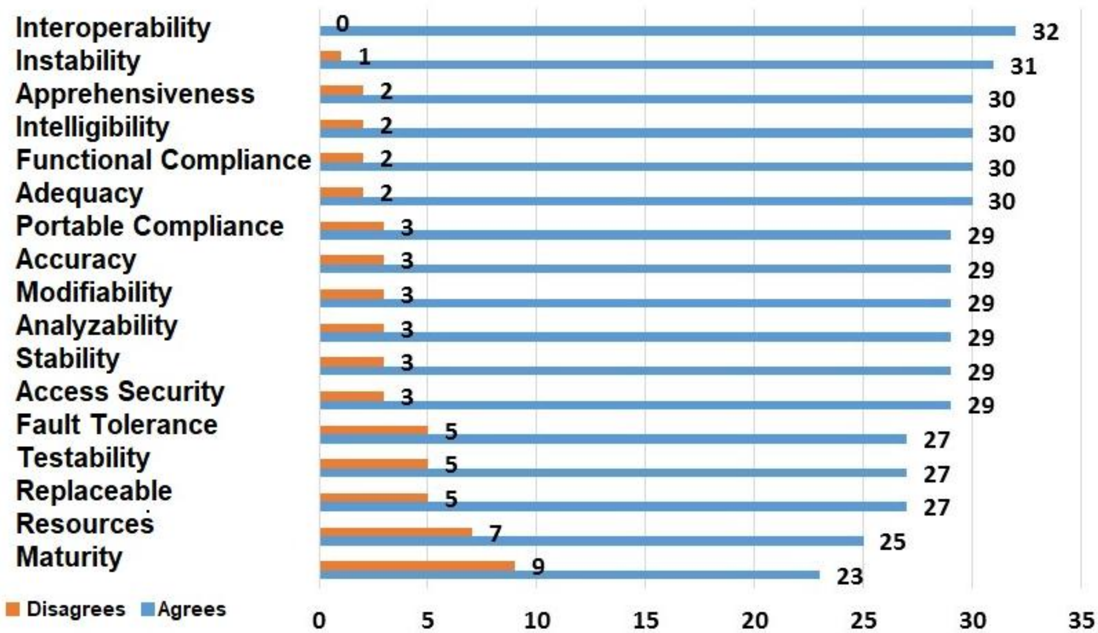


Figure No. 5: Hybrid Quality Dimension (Source: Prepared by the author)

The validation of the three-dimensional model was performed separately for each Quality Dimension. The dimensions of Software Quality, Quality of Use and Hybrid Quality were validated through questionnaires and Pedagogical Quality through interviews. For the Software Quality questionnaire, 32 valid responses (out of 35) were obtained from Software Engineering specialists and 45 (out of 53) from Usability specialists. For Pedagogical Quality, 4 specialists in education and pedagogy were interviewed through 3 virtual meetings lasting 90 minutes each. The Hybrid Quality dimension was analyzed together with the Software Quality and Quality of Use questionnaires and included in the interviews with specialists in education and pedagogy. Altogether, three questionnaires on Software Quality and eight on Quality of Use were discarded, since the training of respondents (Education, Teaching, Physics, Economics, Biological Sciences, Law, Civil Engineering and Physical Education) were not related to Computing.

Of the respondents who are specialists in Software Engineering, 60.0% had a complete doctorate, 25.7% an incomplete doctorate, 5.7% a complete master's degree and 8.6% an incomplete master's degree. Of these, 56.2% are dedicated to teaching, teaching disciplines of Software Engineering, Computing, Requirements, Software Quality, Educational Games, Informatics in Education and Software Evaluation, 21.8% work in the labor market as a software engineer or analyst, 12.5% are software developers. The remaining 9.3% are postgraduate students at masters or doctoral levels.

The amount of agreement and disagreement among Software Engineering specialists follows the distribution shown in Graph 1. Considering a rate of at least 25% of disagreement among respondents (4.8 disagreements), 5 criteria were mentioned as capable of improvement in terms of its nomenclature and description: Fault Tolerance (5), Testability (5), Substitutability (7), Resources (7) and Maturity (9).



Graph 1. Software Quality: Amount of agreement and disagreement

Among the improvement proposals provided by the participants are Fault Tolerance (5 suggestions): (i) “recover from failures”, (ii) “ability to avoid failures” and maintain a specified level of performance in cases of defects in the software or violation of its specified interface”, (iii) “the system's ability to continue operating even in the event of a failure”, (iv) “the ability to mitigate failures and maintain a level of performance specified in cases of defects in the software or violation of its specified interface”, (v) “resilience and resilience when failures occur”. New description: Ability to mitigate failures and maintain a specified level of performance in cases of defects in the software or violation of its specified interface.

Maturity (5 suggestions): (i) “stability”, (ii) “resilience”, (iii) “failure prevention maturity”, (iv) “ability to avoid defects due to software failures”, (v) “ condition acquired by software after several development cycles”. New description: Ability to avoid defects due to software failures.

Testability (4 suggestions): (i) “the software's ability to be tested”, (ii) “the ability to test the modified system, as much as the new features as well as those not directly affected by the modification", (iii) "ability to allow the software, when developed or modified, to be validated ", (iv)" ability to allow the software, when modified, to be verified and validated ". New description: Ability to test the modified system, as much as the new functionalities as well as those not directly affected by the modification.

Substitutability (4 suggestions): (i) "adaptability", (ii) "ability to be comprehensive enough to replace another specified software product, with the same purpose and in the same environment", (iii) "ability to be replaced by another specified software product, with the same purpose, and in the same environment", (iv) "ability to adapt to changes". New description: Ability to be replaced by another specified software product, with the same purpose, and in the same environment.

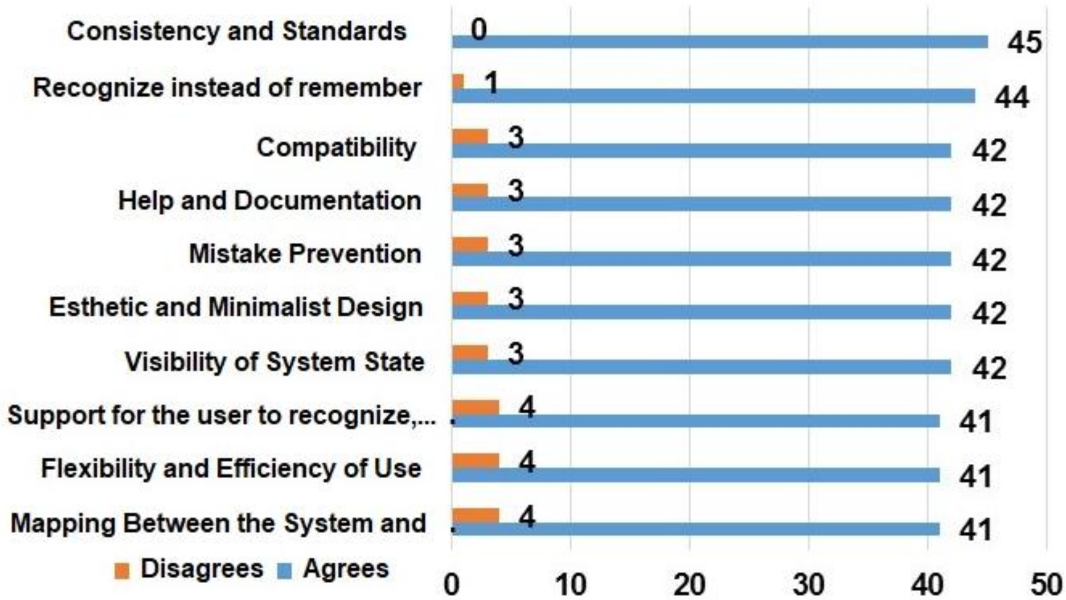
Resources (4 suggestions): (i) "use of resources", (ii) "functionalities", (iii) "ability to use types and quantities of resources within the agreed limits for the software to perform its functions under established conditions", (iv) "technologies, frameworks, libraries, all the support for the proper functioning of the software in question". New nomenclature and description: Use of Resources - Ability to use types and quantities of resources (technologies, frameworks, libraries) within the agreed limits for the software to perform its functions under established conditions.

Even for criteria with a disagreement rate below 15%, some comments were made, for: Suitability: (i) "suitability within a context", (ii) "ability to provide an appropriate set of specific functions for tasks and objectives certain users"; Portable Compliance: (i) "portability"; Accuracy: "ability to provide, with the necessary degree of precision, correct results and effects, or, following pre-established agreements"; Modifiability: (i) "maintainability"; Analyzability: "ability to diagnose deficiencies or causes of software failures"; Stability: (i) "continuity", (ii) "continuous ability to provide service through eventual events", (iii) "ability to respect the projected requirements"; Access Security: (i) "data and access security", (ii) "ability to protect information and data, so that unauthorized persons or systems cannot read or modify them and that the access is not denied access to authorized persons or systems ". However, due to the low rate of disagreement, these criteria have not changed in their nomenclatures or descriptions.

Among respondents who are specialists in Usability Engineering, 42.6% are PhDs, 24.1% have incomplete doctorates, 18.5% have master's degrees, 9.3% have incomplete masters degrees and 5.6% have degrees in Computer Science and related areas. These act as higher education teachers

(37.7%), specialists in educational technologies (35.5%), in User Experience (UX) and Human-Computer Interaction (IHC) (8.8%), software developers (8.8%), and graduate students (4.4%).

The amount of agreement and disagreement of the specialists in Usability Engineering follows the distribution shown in Graph 2. Considering a rate of at least 6.5% of disagreement among respondents (2.92 disagreements), 7 criteria were mentioned as subject to improvement in terms of its description: Compatibility (3), Error prevention (3), Aesthetic and minimalist design (3), Visibility of the system state (3), Support for the user to recognize, diagnose and recover errors (4), Flexibility and efficiency of use (4) and Mapping between the system and the real world (4).



Graph 2. Quality of Use: Quantitative of agreement and disagreement

Among the improvement proposals provided by the participants, there are Support for the user to recognize, diagnose and recover errors (3 suggestions): (i) “validation”, (ii) “Error recovery”, (iii) “to have mechanisms that allow the identification of errors and, when they occur, that favor their correction”. Flexibility and efficiency of use (3 suggestions): (i) “customization”, (ii) “having mechanisms for configuring and personalizing the interface and interaction”, (iii) “having mechanisms for configuring and personalizing the interface and the interaction, and the efficient and effective use related to the user experience”.

Compatibility (2 suggestions): (i) “cognitive leveling”, (ii) “user compatibility”. Error prevention (2 suggestions): (i) "error validation", (ii) "have mechanisms to detect and prevent data entry errors, commands, possible actions with non-recoverable consequences". Aesthetic and minimalist design (2 suggestions): (i) "minimalist design", (ii) "aesthetic design". Visibility of the system status (2 suggestions): (i) “quick feedback”, (ii) “status visibility”.

Mapping between the system and the real world (1 suggestion): (i) “organization between human-machine interactions, according to the user's experiences, using terminologies familiar to the user”. A high level of agreement is perceived for this dimension of taxonomy; therefore, these criteria did not change in their nomenclatures or descriptions.

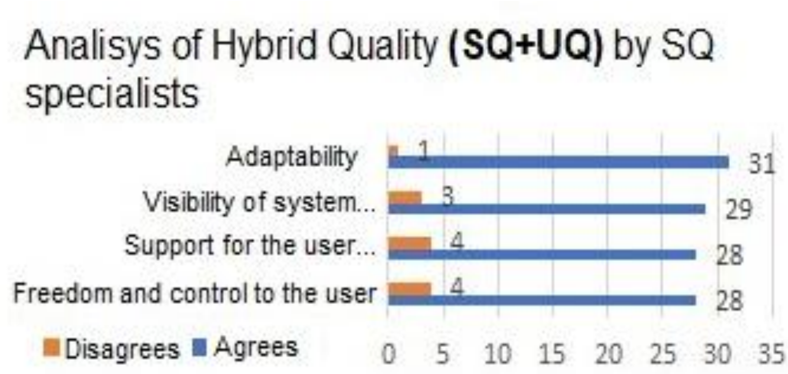
The analysis of the Pedagogical Quality dimension was performed through semi-structured interviews (virtual meetings) with the participation of 4 educators (1 specialist, 1 doctoral student, 2 doctors) who act as teachers (2) and school supervisors (2). The analysis of the educators regarding the nomenclature of the categories and their disposition in the model resulted in some suggestions and modifications; these highlighted that the category Motivational and Responsive Resources should be renamed to “Elaboration of Knowledge” and the order of the categories in the taxonomy should be: previous knowledge → pedagogical foundations and objectives → pedagogical content → relevance of content → epistemological clarity → knowledge elaboration.

For the analysis of criteria nomenclatures, experts considered the criterion/description relationship and its disposition in the categories. During the discussions, they analyzed the agreement of each criterion and the cohesion and coherence of these criteria about the category and description. Given this, they indicated several changes in nomenclatures, the indications of changes were: Pedagogical support guide → “Pedagogical support guide for teachers”; It identifies the pedagogical objectives present in the same → “identification of the pedagogical objectives”; It has different degrees of content complexity in activities → “level of content complexity in activities”; It presents an interdisciplinary approach → “specific or interdisciplinary approach”; It favors group work, but it can also be used individually → “group and or individual work”; Adequacy of the RDA to the content → “Adequacy of the RDA to the content worked”; and, Content correction → “content analysis”.

The criteria descriptions were all modified by them, as they disagreed with all descriptions (agreement, cohesion, and coherence), which were presented for the criteria of this dimension;

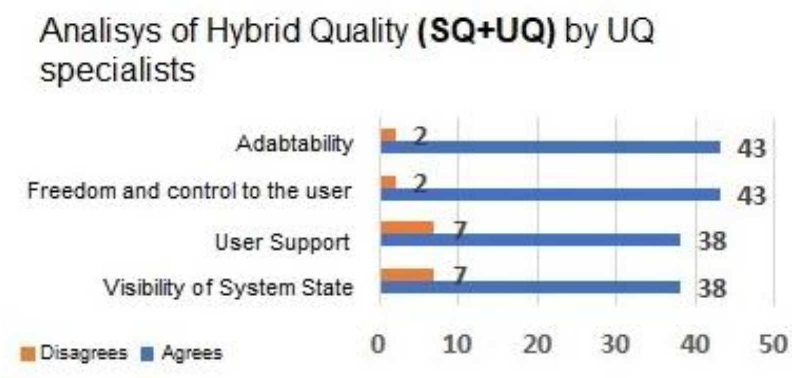
the disagreements also occurred after the changes made in the nomenclatures of the criteria and the allocation of these within the categories.

The profile of the respondents of the Hybrid Dimension follows the same distribution of the software quality dimension presented above. The amount of agreement and disagreement by the specialists in Software Quality and Quality of Use for the hybrid dimension follows the distribution present in Graph 3 and Graph 4. Considering a rate of at least 6% of disagreement among respondents (4.62 disagreements), 3 criteria were mentioned as subject to improvement in terms of their nomenclature and description: Visibility of the system state (5), Support for the user to recognize, diagnose and recover errors (11) and Freedom and control for the user (11).



Graph 3. Hybrid Quality: Amount of agreement and disagreement by experts in Software Quality

Among the improvement proposals provided by participants in Software Quality for the hybrid dimension, there are Freedom and control to the user (3 suggestions): (i) “Freedom and control”, (ii) “configurability”, (iii) “the user's ability to configure the system according to their taste of use and need”. Visibility of the system status (2 suggestions): (i) “audibility”, (ii) “dependability”. Support for the user to recognize, diagnose and recover from errors (1 suggestion): (i) “support to recognize, diagnose and recover from failures”.



Graph 4. Hybrid Quality: Amount of agreement and disagreement by the Quality of Use Specialists

Among the improvement proposals provided by the Quality of Use participants for the hybrid dimension, there are: Visibility of the system status (3 suggestions): (i) "Extra information for the user", (ii) "system or quality of service report", (iii) "have immediate feedback on the system's responses to the user's actions". Support for the user to recognize, diagnose and recover errors (2 suggestions): (i) "error reports", (ii) "immediate feedback", (iii) "have mechanisms to identify errors and, when they occur, that favor their correction". Freedom and control to the user (1 suggestion): (i) "task management".

CONCLUSION:

Considering the importance of evaluating RDAs and understanding that there are challenges regarding the diversity of approaches available in the literature for this purpose, the present work presented as a result a model for three-dimensional evaluation based on the analysis of approaches and quality models.

The model was conceived within research carried out over three years and considered 724 evaluation criteria, extracted from 26 approaches. To favor the multidimensionality in the evaluation of RDAs, the model developed presents a meta-organization that contemplates three dimensions of quality, plus one related to hybrid criteria, 16 categories and 49 criteria, as well as it defines the profile of each evaluator with each dimension, decreasing the effort curve required in this process.

The validation of the model by the specialists impacted some changes in the representation, in the categories, in the nomenclatures and descriptions of the criteria. The highest degree of change was in the dimension of PQ (33%) and the smallest in the dimensions of UQ (0) and HQ (0). The SQ dimension had 29.4% changes and the overall impact of changes in the taxonomy as a whole was corresponding to 12.2%.

As a limitation of this research, it is understood that the study was developed by only one student; however, to minimize the limitation, the study was accompanied by 2 teachers with experience in the areas of IHC, Software Quality, Quality of Use and Evaluation of RDAs. Another limitation was related to the evaluation of the PQ dimension performed by only 4 specialists; to minimize this impact, the interviews were carried out jointly with the 4 educators at the same time.

To continue this research work, we intend to develop a guideline, which guides the evaluation of RDAs, composed of an evaluation scale for each of the criteria that generates a final evaluation score and an application method (step by step) for guide researchers in the planning, execution, analysis and evaluation of RDAs. Also, new interviews will be conducted with specialists in education and pedagogy and a new analysis of the model.

ACKNOWLEDGMENT:

This work was supported by the National Council for Scientific and Technological Development (CNPq) and the Informatics Center of the Federal University of Pernambuco.

REFERENCES:

[1] Goulart, Jorge L. J., Oliveira, Fabiano S., Pinto, Paulo E. D. Avaliacão Sistemática de Eficácia na Aprendizagem de Algoritmos com o uso do Tupy Online. *In: Anais do XXVII Workshop sobre Educação em Computação*. Porto Alegre: SBC, 2019, 218-232.

[2] Barbosa, Rafael S. and de Souza, Ricardo A. C. Innovation Indicators for Educational Softwares. In: *14th Iberian Conference on Information Systems and Technologies (CISTI)*, 2019, 1-6.

[3] Tchounikine, Pierri. Computer science and educational software design: A resource for multidisciplinary work in technology enhanced learning. Springer Science & Business Media: France, 2011.

[4] Medeiros, Angélica F. YPEDUC: Uma adaptação de Metodologia Ágil para o Desenvolvimento de Software Educativo. *In Brazilian Symposium on Computers in Education*, 2019; 30, 1, 379-389.

[5] Brito Junior, Ozonias de Oliveira and Aguiar, Yuska Paola Costa. Taxonomia de Critérios para Avaliação de Software Educativo – TaCASE. *In: Brazilian Symposium on Computers in Education*. 2018; 29, 1: 298-308.

[6] Da Silva, Ana Cristina Barbosa and De França, Rozelma Soares. Avaliação de software educativo na formação docente: articulando teoria e prática. *In: Anais do Workshop de Informática na Escola*. 2014; 20, 1, 106-116.

[7] Petri, Giani and von Wangenheim, Christiane Gresse. How games for computing education are evaluated? A systematic literature review. *In: Computers & education*, 2017, 107, 68-90.

[8] Stéphanie Coomans and Gilberto. S. Lacerda. PETESE, a Pedagogical Ergonomic Tool for Educational Software Evaluation. *In Procedia Manufacturing*, 2015, 3, 5881-5888.

[9] Girardi, Tatiana de Assis, Girardi, Daniel and Marques, Jefferson Luiz Brum. O Uso de um Simulador para o Ensino de Ventilação Mecânica. *In: Revista Brasileira de Informática na Educação*, 2020, 28, 297-319.
[10] Peixoto, Daniela C. C., Resende, Rodolfo. F. and Pádua, Clarindo. I. P. S. Evaluating Software Engineering Simulation Games: The UGALCO framework. *In: IEEE Frontiers in Education Conference (FIE)*, Madrid, Spain, 2014, 1-9.
[11] Brito Junior, Ozonias de Oliveira, Aguiar, Yuska Paola Costa, Tavares, Tatiana Aires. Abordagens para Avaliação de Softwares Educativos e sua Coerência com os Modelos de Qualidade de Software. *In: Brazilian Symposium on Computers in Education*. 2016; 27, 1: 270-280.
[12] Oliveira, Juan and Prata, Wilson. Pesquisa de UX em jogos de Realidade Virtual: desafios e lições aprendidas. *In: Anais Estendidos do XVII Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais*. Porto Alegre: Sociedade Brasileira de Computação, 2018.
[13] Valle, Pedro Henrique Dias, Vilela, Ricardo Ferreira, Júnior, Paulo A. P. and Inocêncio, Ana Carolina Gondim. Hedeg - heurísticas para avaliação de jogos educacionais digitais. *In: Nuevas Ideas en Informática Educativa - TISE*, 2013, 247-256.
[14] Nunes, Raíza Portilho and Santos, Isadora Mendes. A importância da avaliação ergonômico-pedagógica de aplicativos educacionais e os desafios encontrados no campo da aprendizagem móvel. *In Brazilian Symposium on Computers in Education*, 2018, 29, 1, 41-51.
[15] Lin, Yu Han and de Oliveira, Maria Carolina. Avaliação de Software utilizando MAEP: Uma Análise do Protótipo de um Simulador Imersivo de Realidade Virtual. Anais SULCOMP, 2018, 9-10.
[16] Da Silva, Ana Cristina Barbosa and Gomes, Alex Sandro. Conheça e utilize software educativo: avaliação e planejamento para a educação básica. Recife: Pipa Comunicação; 2015.
[17] Pereira, Wendell, Cardoso Filho, Raimundo, Silva, Williane R., Da Silva, Rafael S., Dantas, Vanessa and Aguiar, Yuska P. C. Validação de uma abordagem combinada para avaliação de Software Educativo: avanços e desafios. *In: Revista Tecnologias na Educação*, 2016, 8, 16, 170-180.
[18] De França, Rozelma S. and Da Silva, Ana Cristina B. Avaliação de softwares educativos para o ensino de língua Portuguesa. *In: Revista Brasileira de Informática na Educação*, 2014, 22, 3, 84-98.
[19] Brito Junior, Ozonias de Oliveira and Aguiar, Y. P. C. Análise de abordagens objetivas para avaliação de softwares educativos. *In: Proceedings of the 13th Brazilian Symposium on Human Factors in Computing Systems*. Foz do Iguaçu-PR, 2014.
[20] Iso/IEC 9126-1. Software engineering – Product quality – Part 1: Quality Model. In: Associação Brasileira de Normas Técnicas. 2003.
[21] Swebok. Guide to the Software Engineering Body of Knowledge. Version Project of the IEEE Computer Society Professional Practices Committee. 2004.
[22] Iso/IEC 9214-1. Ergonomic Requirements for office work with visual display terminals – part 11. 2003.
[23] Nielsen, Jakob. Usability engineering. Morgan Kaufman, San Francisco. 1994.
[24] Ben Shneiderman and Catherine Plaisant. Designing the user interface: strategies for effective human-computer interaction (4.Ed). Addison-Wesley Publishing Company. 2004.
[25] Christian B. and Dominique L. S. Ergonomic criteria for the evaluation of human-computer interfaces. RT-0156, INRIA, 79-162. 1993.
[26] De Oliveira, Celina C. and Da Costa José W. Ambientes informatizados de aprendizagem. Produção e Avaliação de Software Educativo. Campinas: Papirus, 2001.
[27] Da Silva, Ana C. B. Softwares Educativos: Critérios de Avaliação a partir dos Discursos da Interface, da Esfera Comunicativa e do Objeto de Ensino. Tese de Doutorado, Universidade Federal de Pernambuco. 2012.
[28] Cresswell, John W. Research Design: qualitative, quantitative, and mixed methods approaches. (4 ed.). Sage, Los Angeles. 2013.
[29] Easterbrook, Steve, Singer, Janice, Storey, Margaret-Anne and Damian, Daniela. Selecting empirical methods for software engineering research. *In: Guide to Advanced Empirical Software Engineering.* Springer, London, 2008, 285–311.

[30] Hevner, Alan R. A three cycle view of design science research. *In: Scandinavian journal of information systems*, 2007, 19, 2, 87-92.
[31] Wohlin, Claes. Guidelines for snowballing in systematic literature studies and a replication in software engineering. *In Proceedings of the 18th international conference on evaluation and assessment in software engineering*, 2014, 1-10.
[32] Reeves, Thomas C. Systematic Evaluation Procedures for Interactive Multimedia for Education and Training. *In Multimedia Computing: preparing for the 21 st century*. Harrisburg, 1994, 472-505.
[33] Campos, Gilda H. B. Metodologia para avaliação da qualidade de software educacional. Diretrizes para desenvolvedores e usuários. Coppe, Rio de Janeiro. 1994.
[34] Muchielli, Alex. L`enseignement par ordinateur. Presses Universitaires, Paris. 1987.
[35] Nesbit, John C., Belfer, Karen and Leacock, Tracey. Learning object review instrument (LORI). *In: E-learning research and assessment network*, 2003, 33-68.
[36] Gamez, Luciano. Técnica de inspeção de conformidade ergonômica de software educacional - TICESE. Dissertação de Mestrado submetida à Universidade do Minho. 1998.
[37] Rocha, Ana Rocha and Campos, Gilda. Avaliação da Qualidade de Software Educacional. Em aberto, Brasília, 1993.
[39] Atayde, Ana Paula R., Teixeira, Adla B. M. and Padua, e Clarindo I. P. S. Metodologia de Avaliação de Qualidade de Software Educacional Infantil – MAQSEI. *In: Anais do XIV Simpósio Brasileiro de Informática na Educação*, Rio de Janeiro, 2003, 1, 1, 356-365.
[39] Gladcheff, Ana Paula, Zuffi, Edna M. and Silva, Dilma M. Um Instrumento para Avaliação da Qualidade de Softwares Educacionais de Matemática para o Ensino Fundamental. *In Anais do XII Simpósio Brasileiro de Informática na Educação (SBIE) e VII Workshop de Informática na Escola (WIE)*, Fortaleza-CE, 2001, 118-128.
[40] Savi, Rafael, von Wangenheim, Christiane Gresse, Ulbricht, Vania and Vanzin, Tarcisio. Proposta de um modelo de avaliação de jogos educacionais. *In: Revista Novas Tecnologias da Educação (RENOTE)*, 2010, 8, 3, 1-12.
[41] Vieira, Fábia M. S. Avaliação de Software Educativo: Reflexões para uma Análise Criteriosa. 12/04/2020. Available from: https://edutec.net/Textos/Alia/MISC/edmagali2.
[42] Silva, Cassandra Ribeiro de Oliveira. MAEP: um método ergopedagógico interativo de avaliação para produtos educacionais informatizados. Tese de Doutorado, Universidade Federal de Santa Catarina, 2002.
[43] Rodrigues, Geovanna C. F. S. Instrumento para avaliação de jogos eletrônicos educativos do Ensino Fundamental I. Dissertação de Mestrado, Universidade Federal da Paraíba, 2014.
[44] Webber, Carine, Boff, Elisa and Bono, Fernanda. Ferramenta Especialista para Avaliação de Software Educacional. *In Anais do XX Simpósio Brasileiro de Informática na Educação*, Florianópolis-SC, 2009, 1, 1, 9-18.
[45] Coskun, Zeynep N., Adıguzel, Tufan and Çatak, Guven. Acoustic labyrinth: validation of a game–based heart auscultation educational tool. *In: World Journal on Educational Technology*, 2019, 11, 4, 245-256.
[46] Petri, Giani, Calderón, Alejandro, von Wangenheim, Christiane G., Borgatto, Adriano and Ruiz, Mercedes. Games for Teaching Software Project Management: An Analysis of the Benefits of Digital and Non-Digital Games. *In Jornal of Universal Computer Science*, 2018, 24, 10, 1424-1451.
[47] Abderrahim, El Mhouti, Mohamed, Erradi and Azeddine, Nasseh. An Evaluation Model of Digital Educational Resources. *In: International Journal of Emerging Technologies in Learning*, 2013, 8, 2, 29-35.
[48] Fang, Haiguang. Modeling and analysis for educational software quality hierarchy triangle. *In: IEEE Seventh International Conference on Web-based Learning*, 2008, 14-18.
[49] Bednarik, Roman, Gerdt, Petri, Miraftabi, Ramin and Tukiainen, Markku. Development of the TUP model-Evaluating educational software. *In: IEEE International Conference on Advanced Learning Technologies*, 2004, 699-701.
[50] Liu, Yuqing and Yang, Liang. The digital learning resource design art and its evaluation. In: IEEE International Conference on Networking and Digital Society, 2010, 2, 331-334.

[51] Besio, Serenella, Laudanna, Elena, Potenza, Francesca, Ferlino, Lucia and Occhionero, Federico. Accessibility of educational software: from evaluation to design guidelines. *In: International Conference on Computers for Handicapped Persons.* Springer, Berlin, Heidelberg. 2008, 518-525.

[52] Tergan, Sigmar-Olaf. Checklists for the evaluation of educational software: critical review and prospects. *In: Innovations in education and training international*, 1998, 35, 1, 9-20.

[53] Sousa, Vanessa E. C., Lopes, Marcos V. O., Ferreira, Gabriele L., Diniz, Camila M., Froes Nathaly B. M. and Sobreira, Bianca A. The construction and evaluation of new educational software for nursing diagnoses: a randomized controlled trial. *In: Nurse education today*, 2016, 36, 221-229.

[54] Yussof, Rahmah L., Paris, Tengku Nazatul Shima Tengku, Abas, Hafiza and Zaman, Halimah Badioze. Mixed Usability Evaluation During the Development Cycle of “MEL-SindD”. *In: Procedia-Social and Behavioral Sciences*, 2013, 105, 162-170.

[55] Secken, Nilgun and Kunduz, Nazan. An evaluation of the educational software about precipitation titrations. *In: Procedia-Social and Behavioral Sciences*, 2013, 106, 387-393.

[56] Squires, David and Preece, Jenny. Usability and learning: evaluating the potential of educational software. In: Computers & education, 1996, 27, 1, 15-22.

 *Author -1*	***Ozonias de Oliveira Brito Junior – Corresponding Author*** *PhD Student Federal University of Pernambuco - Center of Informatics. Recife – PE, Brazil.*
 *Author -2*	***Yuska Paola Costa Aguiar*** *PhD, professor Federal University of Paraiba - Department of Exact Sciences. Rio Tinto – PB, Brazil.*
 *Author -3*	***Hermano Perrelli de Moura*** *PhD, professor Federal University of Pernambuco - Center of Informatics. Recife – PE, Brazil.*