*IOSR Journal Of Humanities And Social Science (IOSR-JHSS)*
*Volume 29, Issue 4, Series 3 (April, 2024) 17-24*
*e-ISSN: 2279-0837, p-ISSN: 2279-0845.*
*www.iosrjournals.org*

# Large-Scale Evaluation Perspectives And Reflections Of Mathematics Teachers

Marcelo Gomes Medeiros De Macedo, Caroline Haussman Dos Santos[1], Francisco Antonio Nascimento [2], Andeson Carlos Santos Moaris[3], Lucas Silva Dos Santos [4], Samuel Soares De Souza Santos [5]

*(Universidade Federal Do Rio Grande Do Norte (Ufrn)– Brasil)*
*[1]((Universidade Federal Do Rio De Janeiro (Ufrj) – Brasil)*
*[2](Universidade Federal De Juiz De For A (Ufjf) – Brasil)*
*[3](Universidade Federal Do Rio Grande Do Norte – (Ufrn), Brasil)*
*[4](Universidade Cidade De São Paulo - Brasil)*
*[5](Instituto Federal De Ciências E Tecnologias Do Amazonas) (Ifam)-, Brasil)*

***Abstract:***
***Background:*** *The study analyzes the perceptions of mathematics teachers on the elaboration of items for large-scale assessments. It emphasizes the importance of Item Response Theory (IRT) in the process.*
***Materials and Methods****: It uses a qualitative approach with a focus on content analysis, based on data collected in discussion forums of an extension course.*
***Results:*** *It reveals the need for clear and contextualized items, highlighting the importance of unidimensionality and the challenges faced by teachers in creating these items.*
***Conclusion****: Emphasizes the importance of continuing training for teachers to improve the quality of large-scale assessment items and understanding of assessment processes.*
***Keywords:*** *Assessment. Item development. Teacher training.*

---

Date of Submission: 02-04-2024 Date of Acceptance: 12-04-2024

---

## I. Introdution

Educational evaluation is a subject that has been much discussed in the context of public education policies. In this sense, the extension course "Public Education Policies: evaluation, curriculum and training of mathematics teachers", promoted by the Study and Research Group Weaving Cognitive Learning Networks (G-TERÇO/CNPq/UFC) and held from May to July 2022, was defined as a training action for teachers who teach mathematics in the public school system, contemplating the remote teaching modality with synchronous meetings through videoconferencing platforms and carrying out asynchronous activities in the Teleduc virtual learning environment, making up a total workload of 60 (sixty) hours.

The main purpose of this extension course was to understand public educational assessment policies and their implications for the mathematics curriculum, and to this end it was structured into five modules that addressed issues and themes related to teaching methodology, curriculum and educational assessment: (1) The Fedathi Sequence methodology; (2) The educational curriculum; (3) The National Common Curriculum Base (BNCC) and the Ceará DCRC Referential Curriculum Document; (4) Public assessment policies and the curriculum; and (5) The development of mathematics items and the Item Response Theory (IRT).

Among the studies carried out in this course, we would highlight the approaches to the development of mathematics items and the IRT, which sought to provide course participants with knowledge about item development techniques and the principles that guide the measurement of the results of large-scale external evaluations.

The aim of this article is to analyze the perception of the mathematics teachers who took part in this extension course about the characteristics that make an item good. This is done from a perspective that takes into account some of the IRT designs that involve the item, such as clarity of language, contextualization, nature and the principle of unidimensionality.

The conceptions of educational assessment pointed out by Freitas et al. (2009) in their different levels contribute to our discussions. As well as the studies by Andrade, Tavares and Valle (2000), Andrade, Brandão and Santos (2020), Couto and Prime (2011) and others about the concepts that permeate IRT.

This article is organized into five sections, the first of which is this introduction. In the second section, we reflect on educational assessment in teaching and learning processes, with an emphasis on large-scale assessments, where we briefly discuss the measurement of their results using IRT, including the importance of the item in this context. In the third section, we present the methodology adopted in this research. In the fourth section, we analyze the teachers' statements and contributions about item development, and finally, in the fifth section, we present our final considerations.

## II. Educational Assessment Processes And Large-Scale Assessment

A Assessment is part of the educational phenomenon since, considering the instances of teaching and learning, someone is always learning something in a given context and all of this goes through constant assessment processes.

For Hadji (2017), evaluation is essential to education, inherent and inseparable when conceived as problematization, questioning and reflection on action. He believes that evaluation cannot cease to exist and is necessary in order to think, question and transform pedagogical actions.

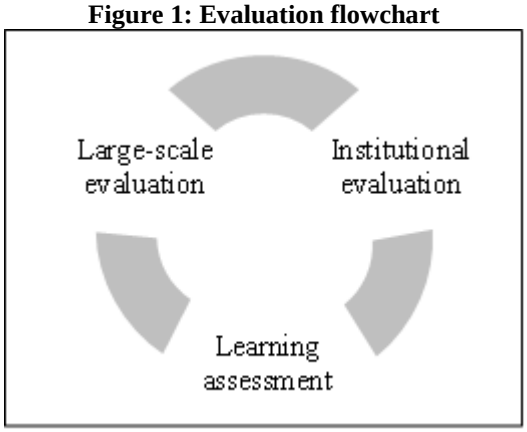
When discussing assessment, it is necessary to realize that its understanding goes far beyond a pedagogical tool to "measure" what the student has actually learned. Assessment should be understood as a powerful lever for increasing student success at school (HADJI, 2017), so assessment should be seen as something that goes beyond just checking, it should be able to understand the situation in which the student finds themselves, providing them with support to discover what they need to improve.

According to Luckesi (2017), assessment in education has often been used as a form of classification, assuming a static and disciplinary role rather than as a means of diagnosis. In the author's view, assessment should be, for the educator, a moment of "reflection".

Freitas et al. (2009) emphasize that evaluation processes can be understood at three levels: (i) learning evaluation; (ii) institutional evaluation and (iii) large-scale evaluation.

Figure 1 shows a flowchart that interrelates these three levels of evaluation.

**Figure 1: Evaluation flowchart**




Source: Adapted from Freitas *et al.* (2009)

Learning assessment involves the teacher-student relationship in the classroom and is based on the definition of certain concepts, including the pedagogical process and the organization of pedagogical work. In this way, the first concept includes assessment, but in this process, Freitas et al. (2009) draw attention to the error of textbooks that place assessment as a formal activity that takes place at the end of the teaching process. As an alternative to this linear view, two main interconnected nuclei or axes are presented, of a dynamic and contradictory nature: objectives/assessment and content/methods. The second concept refers to the organization of pedagogical work, which overlaps at two levels: the classroom and the school, a duplication that allows the authors to argue about learning assessment and institutional assessment, whose focus is on the teacher-student relationship and the school's political-pedagogical project.

Learning assessment becomes meaningful if it is combined with the school's pedagogical project. It supports educational processes and aims to help improve student learning (LUCKESI, 2017).

In turn, the institutional evaluation of a school is a process in which all its actors are involved, with the need to seek an appropriate level of improvement based on the challenges facing the school. While large-scale assessment is external, institutional assessment is internal and under the control of the school, with learning assessment being mainly the responsibility of the classroom teacher. Although these processes are multiple and integrated, each one has its own main protagonist (FREITAS et al., 2009).

It's important to note that not only teachers need to be reflective, but the whole school, in which, in addition to the teachers themselves, staff, students, managers, parents etc. are also involved. Thinking about institutional evaluation therefore means re-evaluating the meaning of the participation of different actors in the life and destiny of schools. In this sense, Freitas et al. (2009) reaffirm that institutional evaluation is a process that involves all of its actors in order to negotiate appropriate levels of improvement based on the specific problems encountered.

Large-scale evaluation, in turn, is a tool for global monitoring of educational networks (network evaluation), making it possible to trace the historical series of system performance, allowing trends to be verified over time in order to reorient public policies. When carried out using an appropriate methodology, it can provide important information on student performance, data on teachers, working conditions and the functioning of schools in the network in Brazil, and is practiced mainly at federal and state level (FREITAS et al., 2009).

According to Freitas et al. (2009), these evaluations would be more effective if they were planned and carried out at municipal level by municipal education councils, which would act as regulators of the evaluation processes of basic education networks, the governing council for the evaluation of the network made up of representatives of the public, administration and school staff, including the parents of the students. This board is responsible for formulating the concepts that will guide the evaluation and operationalizing the evaluation process. Its first action is to delegate to the technical team the development of a reference matrix to create the assessment to be carried out in the network. The second stage is to develop the tests that the students will take. At the same time as developing the assessment instruments, the directorate must set in motion the process of identifying possible areas for data collection.

In this context, it is important to stress that assessment can be seen as a didactic strategy that recognizes the conjectures and hypotheses formulated by the students, the constructive mistakes they make when solving tasks and, in general, their previously acquired knowledge. All of this facilitates the teacher's mediation in teaching and learning processes, as it allows the teaching strategy to be tailored to the students' learning opportunities and the complexity of the object of knowledge.

For Boggino (2009), learning processes should lead students to build and reconstruct knowledge. It is worth remembering that learning means giving new meaning to knowledge. Therefore, it is only possible to interpret reality based on each student's learning possibilities, possibilities given by their cognitive structure, knowledge, values, belief system, etc. Consequently, teaching always involves assessing students' knowledge and proposing relevant strategies, so that students can increasingly restructure and resignify schemes and knowledge and thus reduce the distance that separates them from new curricular knowledge.

Given these different conceptions and levels of assessment, we will focus our studies on large-scale assessments and their models for measuring results, such as the Item Response Theory (IRT) and the elements that make up their assessment instruments, especially the items.

**A brief reflection on the item and Item Response Theory (IRT)**

In order to prepare the standardized tests carried out by large-scale assessments, reference matrices are used. These are "cut-outs" of teaching curriculum proposals, drawn up by experts in assessment and in specific areas of knowledge, containing a set of skills that are expected of students. The skills indicated in the reference matrices serve as the basis for drawing up the items, which are the questions that make up a large-scale assessment. In general, items consist of a statement, support, command and alternative answers, which can be wrong, called distractors, or correct, called a template. The items are designed to assess only one skill, seeking to address a single dimension of knowledge (ANDRADE; BRANDÃO; SANTOS, 2020).

The standardized tests in large-scale assessments use predominantly multiple-choice items, which is one of their limitations. In addition, the fact that they do not cover the entire curriculum accentuates the restrictions of these assessments.

In this context, the IRT helps to measure the results of these large-scale assessments. It consists of a set of mathematical models that seek to represent the probability of an individual getting a particular item right as a function of the respondent's ability (or abilities). This relationship is always expressed in such a way that the greater the ability, the greater the probability of getting the item right.

According to the studies by Andrade, Tavares and Valle (2000), Andrade, Brandão and Santos (2020) and Couto and Prime (2011), the different models proposed in the literature basically depend on three factors: (i) the nature of the item; (ii) the number of populations involved - just one or more than one; (iii) and the number of latent traits being measured - just one or more than one.

Andrade, Brandão, Santos (2020) point out that, regarding the nature of item (i), we have dichotomous items, which are characterized by being multiple choice, in which there is only one possibility of getting it right, and non-dichotomous items, which are characterized by being open-ended, with a free response, allowing for more than one category of getting it right.

As for the number of populations involved (ii), you can have one or more than one. However, it is important to understand that it is common in the field of educational assessment for the population to be defined by characteristics that may vary according to the purpose of the study. So, for example, students in the 9th year of elementary school and those in the 3rd year of secondary school can be considered different populations. It is also possible to consider different populations, students from the same year, but from different schooling periods, for example, 9th grade students in 2020 and in 2021. Once you understand the concept of population, you can then understand what a group means as a population sample. For example, a group of morning shift students and another group of afternoon shift students can be selected from the population of 5th grade students (ANDRADE; TAVARES; VALLE, 2000).

With regard to the number of latent traits measured (iii), Andrade; Tavares; Valle (2000) argue that you can have one or more than one. A latent trait can be understood as an individual's degree of satisfaction or maturity in relation to skill or expertise. Large assessment models generally use the principle of unidimensionality. In other words, they are made up of items that are supposed to measure a single latent trait. In other words, there should be only one skill responsible for performing the tested item, or at least one skill considered dominant (TAVARES, 2013).

In summary, IRT is an analysis approach based on the item's parameters: (a) discrimination, (b) degree of difficulty and (c) the chance of getting it right by chance. By examining these parameters based on the answers provided by the students in the test, IRT estimates the student's level of ability using mathematical models called Maximum Likelihood Estimators (MLE). This is because the IRT postulates that the higher a subject's ability, the more likely they are to get items right that are related to their learning level or lower. For example, if a student has a certain ability, then they will find it easier to get items right at the same or lower level, and in turn, they will find it harder to get items right at a higher level of difficulty than their ability.

According to Andrade, Tavares and Valle (2000), one of the great advantages of using IRT is that it allows comparisons to be made between the results of individuals from different populations when submitted to tests that have some items in common, and also allows comparisons to be made between individuals from the same population submitted to completely different tests. This is possible because the central element of IRT analysis is the items and not the test as a whole.

In addition, another important advantage is that the IRT makes it possible to identify skills that have already been consolidated and those that are still developing, which makes it possible to identify possible learning gaps. As a result, teachers can adapt their teaching practices to overcome students' difficulties, so that they can reach the desired learning levels.

## III. Methodological Path

From the point of view of its nature, methodologically this research is configured as being of a basic nature, as it collaborates in the elaboration of new and useful knowledge for the advancement of science, without necessarily having a practical application (PRODANOV; FREITAS, 2013).

In order to achieve our objective, we have adopted a descriptive-exploratory study, since its design is very flexible and features both the description of the phenomena observed, emphasizing their causes and relationships with other facts, and the analysis and interpretation of the data collected.

In our analysis, we favored the qualitative approach because it considers that there is a dynamic relationship between the real world and the subject, that is, an inseparable link between the objective world and the subjectivity of the subject that cannot be translated into numbers. The interpretation of phenomena and the attribution of meanings are fundamental to the qualitative research process.

Our research procedure included elements of participant research, since it is characterized by interaction between researchers and members of the situations being investigated. Participatory research, in turn, involves the community in the analysis of its own reality, is developed through interaction between researchers and subjects, with the aim of finding real problems to be debated and studied, in other words, it makes it possible to create, work with and interpret reality, above all using the resources offered by the researched environment (GIL, 2010).

Participatory research, as its name suggests, involves the participation of the research subjects, combining the interrelationship between research and action in a particular field selected by the researcher. According to Gil (2010, p. 31), this type of research "[...] is characterized by the involvement of the researcher and those being researched in the research process". In this way, both the researcher and the participants are cooperatively involved in the research work.

The subjects of this research are the 31 teachers who took part in the extension course. In order not to identify the participants, we have used the following nomenclature to represent them: P1, P2, P3, P4 and so on.

The research design was based on data collection in the extension course discussion forums. As a criterion for participation, we used the fact that the post was related to the principles of item elaboration and characterization and to the designs of the IRT.

To analyze the data, we adopted elements of content analysis (BARDIN, 2016), which includes data processing techniques in qualitative research and has been used to interpret books and other writings since the dawn of man.

Bardin (2016) defines the content analysis approach as a set of communication analysis techniques aimed at obtaining indicators (quantitative or non-quantitative) through systematic and objective procedures for describing the content of messages that allow the inference of knowledge regarding the conditions of production.

This method of analysis takes place in three phases: (1) preliminary analysis; (2) exploration of the material; and (3) treatment of the results.

In the preliminary analysis phase (1) we proceeded to select, organize and study the data collected in the discussion forums. In the material exploration phase (2) we proceeded to code and characterize the data collected in order to define the categories of analysis. In this process we defined the following categories referring to the characteristics of a good item: (i) use of clear language and contextualization; (ii) nature of the item; (iii) principle of unidimensionality; and (iv) difficulties in developing items. In the results processing phase (3), we proceeded with the analysis and interpretation of the data in the light of the theoretical precepts, seeking at all times to analyze the perception of the subjects of this research about the characteristics that make a math item considered good.

## IV. Results And Discussions

In this section, we analyze the contributions of the teachers who took part in the extension course, seeking to interpret their perceptions of the characteristics of a good item from a perspective that takes into account the aims of IRT and involves the processes of item development.

Based on the data collected in the discussion forums, and using the content analysis methodology (BARDIN, 2016), we present below the analysis of each category defined in the material exploration phase.

**Use clear language and contextualization**

The basic element for a good math item is its clarity and objectivity. To achieve this, the item must have simple language that is accessible to students, without the use of far-fetched words or techniques that make it difficult to understand.

In the item, the statement must contain all the information needed to solve it, and may use a support such as an image, table, illustration or other resource. It must also be presented in an attractive and challenging way, encouraging students to seek a solution.

The wording should also provide clear and concise information about the item, which is consistent with the skill and does not contain unnecessary information. The wording should make sense in the context of the item and should be easy to understand. All of this will help to ensure that users have a better understanding of what is required by the item.

On this subject, among the data collected, we highlight the words of teachers P2, P5, P7, P9, P12, P14, P15, P19, P27 and P31, who state that one of the aspects that makes an item good is the use of clear language. For these teachers, an item needs to contain clear information in its wording, consistent with the skill, avoiding unnecessary information in the base text, but remembering that it must contain all the information needed to solve the item.

Teacher P9 explains that the item "should also have language that is easy for students to understand and appropriate for their age group". P14 complements this thought by pointing out that the basic texts should preferably be short, "up to 15 lines" and present a good wording such as the question to be answered, having clarity and objectivity in its message, so it is not interesting to present elements that make it difficult for the student to understand the item.

On this aspect, other subjects commented that:

There are several factors that lead to an item being considered good, it must be planned according to the content being worked on, the teacher must use clear language that doesn't present any doubts and the text must be short, without unnecessary information, because the aim is to assess what the student has learned, or needs to learn more (P2).

An item needs to contain clear information in its command, consistent with the skill/writer being worked on, be contextualized to the student's daily life, remember to maintain the same semantic field when offering distractors and always try not to mislead the student (P15).

A good item, as discussed in this module, should have a good wording as the question to be answered, be clear in its wording, be objective; it is not interesting to present elements that make it difficult for the student to understand the item. The item can also lead the student to reflect on what it proposes. The questions presented in the items should have content that leads the student to imagine and create in order to achieve what they really know (P27).

Among the aspects highlighted, one of the main ones is the use of clear, concise language that is coherent with the skill covered by the item, avoiding unnecessary information and being easy to understand. In addition, the wording should be succinct enough not to take up too much space, but detailed enough to convey all the necessary information. Other factors are also mentioned, such as the importance of using language appropriate to the age group of the student and formulating clear and objective statements. In short, the choice of words and the structure of the statement are fundamental elements in the construction of an effective and fair item, which really assesses what the student knows or needs to learn.

**Nature of the item**

Couto and Primi (2011) emphasize the importance of dichotomous items for measuring respondents' abilities, as such items only require a simple answer, with no room for ambiguity. This allows for a more reliable assessment, as there are no assumptions involved. However, non-dichotomous items can also be useful, as they provide a richer understanding of the subject, allowing for more in-depth analysis and insights. For example, open-ended questions can be used to assess problem-solving skills or creative thinking. The most appropriate approach will depend on the specific nature of the item and what it is intended to measure.

The difficulty of the item should also be taken into account when making decisions about the type of item to use. For example, a multiple choice item may be appropriate for a simple concept, while an item involving argumentative text production may be more suitable for a complex concept.

In this regard, we would like to highlight the statements made by teachers P5, P8, P9, P11, P12, P13, P14, P16 and P29, who consider the nature of the item, highlighting the characteristics of dichotomous items that are more prevalent in large-scale assessment tests.

The teachers emphasized that the purpose of the items is to measure previously defined skills, so that they don't encourage the famous "rote learning" and don't lead students into errors or so-called "pranks". For teacher P9, "the answer alternatives must be plausible, indicating possible paths that the student can use to reach certain answers as correct". Teacher P14 adds that the wording of the item must be clear and objective, without giving rise to dubious interpretations.

On this subject, among the comments made in the discussion forums, we would highlight the following:

The item needs to measure a single learning outcome, and it is right to choose the descriptor and then the basic text, and not try to "fit" an item into a descriptor once it is ready. That's why it's important to do a lot of research into texts and topics in a variety of sources and not just stick to textbooks (P11).

A good item is one that: a) is coherent with the descriptor indicated; b) is clearly and objectively worded, with no room for dubious interpretations; c) presents coherent answers, a template and coherent distractors. I also consider a good item to be one that enables students to reflect on their learning (P13).

[...] we can also highlight its nature, because a good item, like a good assessment, cannot be centered only on mechanical and operative procedures, such as applying formulas or performing arithmetic and/or algebraic operations. A well-designed item should immerse the student in a challenging and meaningful problem situation, so that the student feels motivated to be the protagonist of an investigative action, thereby developing creative and innovative strategies in the search for a solution to the problem (P29).

The teachers pointed out that the items should be comprehensible, objective and free of ambiguity, avoiding the use of mnemonic devices or the application of ready-made rules or formulas. They emphasized that designing items requires practice, dedication and careful study of what is to be assessed, starting with the choice of skill and the basic text that will support the wording. In this way, the aim is to create items that allow the student's abilities to be assessed, while being contextualized and coherent, without leading to errors or misinterpretations, and promoting student reflection and inquiry.

**Principle Of One-Dimensionality**

We know that large-scale assessment models are generally one-dimensional. This means that they are made up of items that are supposed to be measuring a single skill, or at least a skill that is considered dominant (TAVARES, 2013).

In order to have an item that meets the precepts of unidimensionality, there must be strict criteria in its preparation, and for this to happen, it must respect the requirements that will be analyzed, such as being duly related to a particular skill in the reference matrix (COELHO, 2014).

In this respect, teachers P3, P7, P9, P10, P13, P17, P18 and P21 almost all agree on the need for the item to assess a single skill. Teacher P7 points out that "to be considered a good item, one of the main characteristics that must be observed is whether it meets the skill described in a one-dimensional way by assessing only one skill".

In this context, we highlight the following contributions from the subjects investigated:

In order for an item to be considered good, it must meet a number of requirements, including being original, being in line with the reference matrices, assessing a single skill, being appropriate for the grade it is intended for, having clear and precise language, having a perfect structure, as well as having a coherent wording, without addressing negative issues or misleading the student (P3).

The purpose of the items is to measure the student's abilities in relation to a previously defined descriptor. The item needs to be in line with the reference matrices (which are only a small part of the curriculum) (P13)

It must fully correspond to what the descriptor asks for, the questions must be autonomous, unrelated to another question, have simple and direct texts, avoid personal preferences, opinions, value judgments, without religious and political references (P18).

The teachers emphasized the importance of an item assessing a single skill in a one-dimensional way, with plausible alternatives and clear language appropriate to the student's age group. In this way, the skill required by the item must strictly comply with that described in the reference matrix.

Participants also highlighted the importance of having adequate support, properly referenced, well-designed alternatives that don't mislead the student, and the template must be organized in such a way that it doesn't distract the student or mislead them into answering the question. In other words, designing an item requires practice and commitment, as well as in-depth knowledge and dedication.

**Difficulties In Preparing Items**

Preparing a math item, especially one that will be used in large-scale assessments, is not an easy task; it requires careful practice, dedication and a deep understanding of the knowledge being assessed. It is important to ensure that all aspects of the assessment item are worded correctly, so that students understand what is being assessed. In addition, it is necessary to provide adequate feedback to students so that they can understand their performance and the areas in which they need to improve.

Teachers P8, P16, P17, P22, P23, P24 and P25 made many comments emphasizing the difficulties involved in writing an item. Starting from this premise, teacher P8 said that "drafting the item is not a task that requires practice, dedication and careful study of what you want to achieve". P16 adds that "designing an item is not an easy task", reinforced by P17 when he says that "teaching experience is of fundamental importance in order to be able to design items in line with the educational context".

Other teachers also emphasized in their posts the difficulty in preparing items:

[...] many teachers don't know how to prepare an item according to the necessary techniques. Although it's not a simple task, it's important that we, as teachers, try to get to grips with these techniques so that we can understand how our students are assessed (P24).

[...] preparing an item is very difficult, especially when you don't have any practice. Most of the time, and I include myself in this group, it's much more practical to search for ready-made questions on websites or in books and use them in classroom exercises or assessments, than to prepare items. In short, it's a habit that needs to be changed, because practicing making items makes us learn much more (P25).

Teachers' work overload can also lead to a lack of precision in item development. Teaching practice is essential for the development of an item and for teachers to be able to understand how their students are assessed. However, it is necessary to have a working day that makes it possible to reconcile study and planning processes with teaching activities.

The dissemination of knowledge on how to design a good item is essential to guide teachers on the components that make an item suitable and how to design it properly. In this way, training processes are becoming increasingly necessary so that teachers can better understand how to design a good math item.

## V. Final Considerations

The aim of this study was to analyze the perception of mathematics teachers, the subjects of this research, about the characteristics that make an item good. In order to do this, we prioritized our analysis on the IRT designs that involve the item and its elaboration process.

We analyzed 38 posts from the extension course's discussion forums related to module 5 of the course, which deals with the development of mathematics items and the Item Response Theory (IRT).

In the previous sections we discussed the processes that involve educational assessment and its levels, according to Freitas *et al.* (2009), emphasizing large-scale assessments, the principles that permeate the measurement of their results through IRT and the characteristics of the items that make up the instruments of these assessments.

In our analysis of the quality of an item, based on the perceptions of the teachers who took part in this research, we highlighted the following characteristics that make an item considered good: the use of clear and objective language; contextualization with the subject being assessed; being essentially a dichotomous item when

it comes to tests with multiple-choice questions; and the need for the item to assess a single skill or at least one dominant skill, thus complying with the principle of unidimensionality.

In our findings, we also found that, in the perception of the subjects of this research, the process of preparing an item is not simple and requires a lot of dedication and preparation on the part of the teachers.

Finally, we emphasize the importance of initiatives such as the extension course investigated in this research, since it is training actions like this that help teachers expand their knowledge and qualify their pedagogical practices, whether in the preparation of mathematics items or in a better understanding of the assessment processes present in large-scale assessments.

## References


[1] Andrade, Wendel Melo; Brandão, Jorge Carvalho; Santos, Maria José Costa Dos. Analysis Of Mathematics Item Parameters In The Light Of Classical Test Theory (Ctt) And Item Response Theory (Irt). Perspectivas Da Educação Matemática - Inma/Ufms - V. 13, N. 32, 2020.

[2] Andrade, D. F. Tavares, H. R.; Valle, R. C.Item Response Theory: Concepts And Applications. Brazilian Statistical Association. São Paulo: Sinape. 2000.

[3] Bardin, Laurence. Content Analysis. São Paulo: Editions 70. 2016.

[4] Boggino, Norberto. Evaluation As A Teaching Strategy. Evaluating Processes And Results. Revista De Ciências Da Educação. N. 9, 2009.

[5] Coelho, E. C. Item Response Theory: Challenges And Perspectives In Multidisciplinary Exams. Thesis (Doctorate In Science). Federal University Of Paraná. Curitiba, 2014.

[6] Couto, Gleiber; Primi, Ricardo. Item Response Theory (Irt): Elementary Concepts Of Models For Dichotomous Items. Bol. Psicol [Online]. V. 61, N. 134, 2011.

[7] Freitas, L. C. Et Al. Educational Evaluation: Going Against The Grain. Educational Frontiers Collection. Petrópolis: Vozes, 2009.

[8] Gil. How To Design Research Projects. 5. Ed. São Paulo: Atlas, 2010.

[9] Hadji, Charles. Assessment Demythologized. Porto Alegre: Artmed, 2017.

[10] Luckesi, Cipriano Carlos. Evaluation Of School Learning. 10. Ed., São Paulo: Cortez, 2017.

[11] Prodanov, Cleber Cristiano; Freitas, Ernani Cesar De. Metodologia Do Trabalho Científico: Métodos E Técnicas Da Pesquisa E Do Trabalho Acadêmico. 2. Ed. Novo Hamburgo: Feevale, 2013.

[12] Tavares, C. Z. Item Response Theory: A Critical Analysis Of Epistemological Assumptions. Estudos Em Avaliação Educacional, São Paulo, V. 24, N. 54, P. 56-76, Jan./Abr. 2013.