sustainability

MDPI

Article

Towards Digital Transformation of the Validation and Triage Process of Textbooks in the Brazilian Educational Policy

Álvaro Sobrinho [1,*,†], Ig Ibert Bittencourt [2,3,†], Andressa Carvalho Melo da Silveira [4,†], Alan Pedro da Silva [2,†], Diego Dermeval [2,†], Leonardo Brandão Marques [2,†], Nadja Cezar Ianzer Rodrigues [5,†], Ana Carolina Silva e Souza [5,†], Rafael Ferreira [6,7,†] and Seiji Isotani [3,8,†]

1 Computer Science, Federal University of the Agreste of Pernambuco, Garanhuns 55292-270, Brazil
2 Computing Institute, Federal University of Alagoas, Maceió 57072-260, Brazil; ig.ibert@ic.ufal.br (I.I.B.); alanpedro@ic.ufal.br (A.P.d.S.); diego.matos@famed.ufal.br (D.D.); leonardo.marques@cedu.ufal.br (L.B.M.)
3 Harvard Graduate School of Education, Harvard University, Cambridge, MA 02138, USA; seiji_isotani@gse.harvard.edu
4 Electrical Engineering Department, Federal University of Campina Grande, Campina Grande 58429-900, Brazil; andressa.queiroz@ee.ufcg.edu.br
5 Brazilian National Education Development Fund, FNDE, Brasília 70297-400, Brazil; nadja.rodrigues@fnde.gov.br (N.C.I.R.); ana.souza@fnde.gov.br (A.C.S.e.S.)
6 Artificial Intelligence Lab, Federal Rural University of Pernambuco, Recife 52171-900, Brazil; rafael.mello@ufrpe.br
7 Studies and Advanced Systems Center, CESAR School, Recife 50030-220, Brazil
8 Department of Computer Systems, University of São Paulo, São Carlos 13566-590, Brazil
* Correspondence: alvaro.alvares@ufape.edu.br
† These authors contributed equally to this work.

Abstract: One factor that impacts the quality of Brazilian education is the quality of books and other didactic materials freely distributed throughout the country to public schools, thanks to the Brazilian National Textbook Program. The current evaluation process may take at least two years to complete, involving hundreds of people, and the final result may impact the entire educational system. One of the first activities of the process is to validate and triage the editorial quality attributes of textbooks. However, the validation and triage process needs improvement, considering the gradual expansion of the quantity and variety of materials that currently affect it. This generates risks of reduced quality and timely deliveries. This paper provides a comprehensive critical analysis of the validation and triage process based on the Policy Design Arc framework of Harvard's Kennedy School of Government. We identified causes that affect the quality of deliveries and the time required to conclude tasks. We also propose a theory of change for digital transformation, defining strategies to address the causes of problems, outputs, outcomes, and impacts. Therefore, we have gradually implemented our theory of change in the validation and triage process.

Keywords: theory of change; Ishikawa diagram; Policy Design Arc; sustainable development goals

check for updates

Citation: Sobrinho, Á.; Ibert Bittencourt, I.; Carvalho Melo da Silveira, A.; Pedro da Silva, A.; Dermeval, D.; Brandão Marques, L.; Cezar Ianzer Rodrigues, N.; Carolina Silva e Souza, A.; Ferreira, R.; Isotani, S. Towards Digital Transformation of the Validation and Triage Process of Textbooks in the Brazilian Educational Policy. *Sustainability* **2023**, *15*, 5861. https://doi.org/10.3390/su15075861

Academic Editor: Yuto Kitamura

Received: 7 February 2023
Revised: 13 March 2023
Accepted: 26 March 2023
Published: 28 March 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The Brazilian National Textbook Program (Programa Nacional do Livro e do Material Didático—PNLD) is an example of a Brazilian public policy designed to benefit students and teachers in public basic education schools and non-profit organizations [1]. Created in 1985, PNLD comprises a set of planned actions to distribute didactic, pedagogical, and literary books and other educational materials to support academic practice. This public policy helps the Brazilian government improve the quality of books and other materials available for students and teachers and, consequently, improve the teaching–learning process in public schools. The PNLD requires publishers to submit their books and materials through an evaluation process to ensure compliance with quality standards

and pedagogical needs defined in the Brazilian curriculum. This program is recognized as a relevant public policy in Brazil.

The evaluation process of such materials is a legal responsibility of the Brazilian National Education Development Fund (Fundo Nacional de Desenvolvimento da Educação—FNDE). The FNDE's mission is to transfer financial resources and provide technical assistance to guarantee quality education for all. In this sense, the FNDE has Collaborating Centers for Didactic Materials and Support for Educational Practice (Centros Colaboradores em Materiais Didáticos e de Apoio à Prática Educativa—Cepli) so that all stages of the PNLD take place within the standards required by the program's legislation and within the period necessary for the books to be in the school before the beginning of the school year. Therefore, the FNDE relies on partnerships with institutions with the technical capacity to support the execution of this process. Figure 1 provides an overview of the evaluation process of textbooks in Brazilian educational policy. The first activity is the enrollment and validation of textbooks, which is the article's focus. Then, the textbooks validated by the FNDE are pedagogically evaluated by the Ministry of Education (Ministério da Educação—MEC). Once the pedagogical evaluation is concluded, responsibility returns to the FNDE to complete the process through technical qualification, choice (by public school principals and teachers), negotiation, acquisition, and distribution of textbooks. Therefore, the pedagogical evaluation is currently the only activity performed outside the scope of the FNDE.

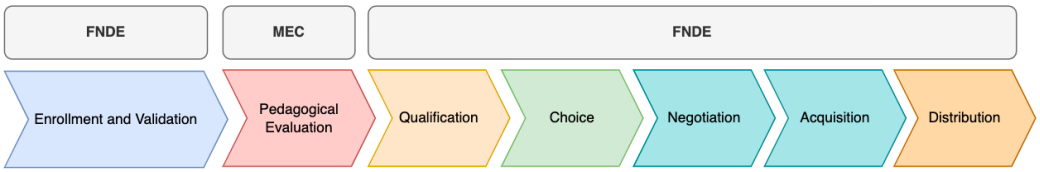

Figure 1. Overview of the complete process for evaluating textbooks in the Brazilian educational policy.

Starting from the second semester of 2021, the Center of Excellence for Social Technologies (Núcleo de Excelência em Tecnologias Sociais—NEES) at the Federal University of Alagoas (Universidade Federal de Alagoas—UFAL) became responsible for the validation and triage process in the PNLD, which motivated the proposal for digital transformation.

However, the PNLD faces challenges, especially with the efficiency and effectiveness of validating and analyzing the editorial quality attributes of books and educational materials. The entire evaluation process can take at least two years to complete, involving hundreds of people, and the final result may not guarantee the quality of the materials. The validation process, which is the screening stage of the PNLD, is currently affected by the gradual expansion of the quantity and variety of materials, resulting in risks of reduced quality in deliveries and increased time to complete tasks. When the validation and triage process performs well (i.e., tasks are conducted effectively and efficiently), the subsequent activities of the evaluation process are positively impacted.

For instance, the validation and analysis of editorial quality attributes are currently manually conducted and managed using spreadsheets, which is an error-prone and time-consuming process. These problems can lead to low-quality analysis and increase the time required to provide materials for students and teachers, compromising the achievement of relevant and effective learning outcomes expected by the Brazilian population.

This article presents results from a research project with a broad scope of studying and applying augmented intelligence in validating and analyzing attributes of teaching materials and digital resources of the PNLD. This research provides a comprehensive critical analysis of the process for validating and analyzing editorial quality attributes of books and didactic materials from PNLD. Therefore, the pedagogical evaluation of textbooks is outside the scope of this study. We examine and discuss the current validation and triage process of textbooks based on the Policy Design Arc framework of Harvard's Kennedy School of Government. We present the causes of problems and their effects by developing a cause–effect diagram (Ishikawa diagram) [2]. Based on the Ishikawa diagram, this paper

presents a theory of change [3] that discusses strategies, outputs, immediate outcomes, intermediate outcomes, the primary outcome, Sustainable Development Goals (SDGs), and impact. Therefore, we have gradually implemented our theory of change in the validation and triage process of the PNLD. This is the first study addressing the digital transformation of the textbooks' validation and triage process in Brazilian educational policy. This article contributes to sustainability in the context of sustainable education because the availability of high-quality digital books is crucial to increase equity in Brazilian public schools, helping achieve relevant SDGs. Furthermore, improving the accessibility of digital textbooks can assist the Brazilian government in advancing equity in Brazilian public schools.

Therefore, this paper is organized as follows. In the Methodology section, we introduce our methodology and the conceptual basis for constructing Ishikawa diagrams, Cohen's Kappa statistic, and the theory of change. Afterward, we present the critical analysis conducted on the process for validation and analysis of editorial quality attributes of textbooks and didactic materials. Then, we discuss this process's recommended strategies, expected outputs, outcomes, and impacts based on the Ishikawa diagram and the developed theory of change.

2. Related Works

The digital transformation of public administration has been discussed in previous research [4–6]. For instance, Scupola and Mergel [7] investigated Denmark's digital strategy formulation and implementation. The authors reviewed the existing literature and concluded that co-production was essential to digital transformation. Giulio and Vecchi [8] discussed the technological changes that guide the structure and strategies of public administrations based on the e-government reform in the Italian public sector. Filgueiras et al. [4] discussed the digital transformation in the Brazilian context, focusing on public service delivery.

Previous research has also investigated the Brazilian national textbook program, such as [1,9–13]. Höfling [9] criticizes, from a political point of view, the focus of the PNLD on specific private publishing groups during the decision-making process. However, nowadays, it is possible to observe that the current evaluation process reaches a broader range of publishers submitting textbooks for analysis. Zambon and Terrazzan [1] analyze the criteria used by public schools to choose textbooks. Thus, the authors interviewed members of public school management teams and state that two factors influenced the choice of textbooks by public schools:

1. actions developed by publishers rather than guidelines from the FNDE;
2. opinions of teachers collected during brief meetings.

Bianco [10] analyzed nutrition education textbooks approved in the PNLD 2014. The authors identified flaws in the imagery content of nine collections of science textbooks approved by the PNLD. In addition, Souza and Rego [11] analyzed images of science and physics textbooks, relating them to the 2018 PNLD guidelines. The authors state that only some PNLD guidelines were properly used while analyzing the images presented in the selected textbooks. Manoel et al. [12] studied the mathematics textbooks approved by the PNLD in 2015, focusing on financial mathematics. They discuss the impacts of textbooks on society.

Albuquerque and Ferreira [13] analyzed the changes in the evaluation criteria of textbooks related to PNLD. The authors interviewed school teachers and analyzed the PNLD guidelines for 2007, 2010, and 2013. They concluded that the PNLD has contributed to the availability of materials that assist teachers during literacy practices.

Nevertheless, to date, the studies still need to address the process of validating and analyzing editorial quality attributes of didactic, pedagogical, and literary books and other materials. Thus, our study is the first one addressing this topic and it is guided by the following research question: *What are the current causes of problems related to the validation and triage process of textbooks and the possible solutions to address the problems?*

3. Methodology

This research considers the combined usage of the Ishikawa diagram and the theory of change to analyze the validation and triage process of textbooks in Brazilian educational policy. Thus, we applied the Policy Design Arc framework [14]. The Policy Design Arc is a framework for strategic thinking created at Harvard's Kennedy School of Government. When applying such a framework, it is necessary to consider values, problems, technical, political, and organizational issues, and recommendations.

The methodology, defined based on the Policy Design Arc framework, is depicted in Figure 2. This study defines values as the first step in investigating the current validation and triage process of textbooks in Brazilian educational policy. This analysis was guided by semi-structured interviews with stakeholders (e.g., quality managers) and was fundamental in identifying causes and effects. The output of the problem step is an Ishikawa diagram for cause–effect analysis, supported by evidence generated using interviews with stakeholders, internal documents, and Cohen's Kappa statistic results. The Ishikawa diagram is the basis for analyzing possible technical solutions and political and organizational issues. The analysis steps drive the theory of change, proposing the digital transformation of the validation and triage process of textbooks in Brazilian educational policy.

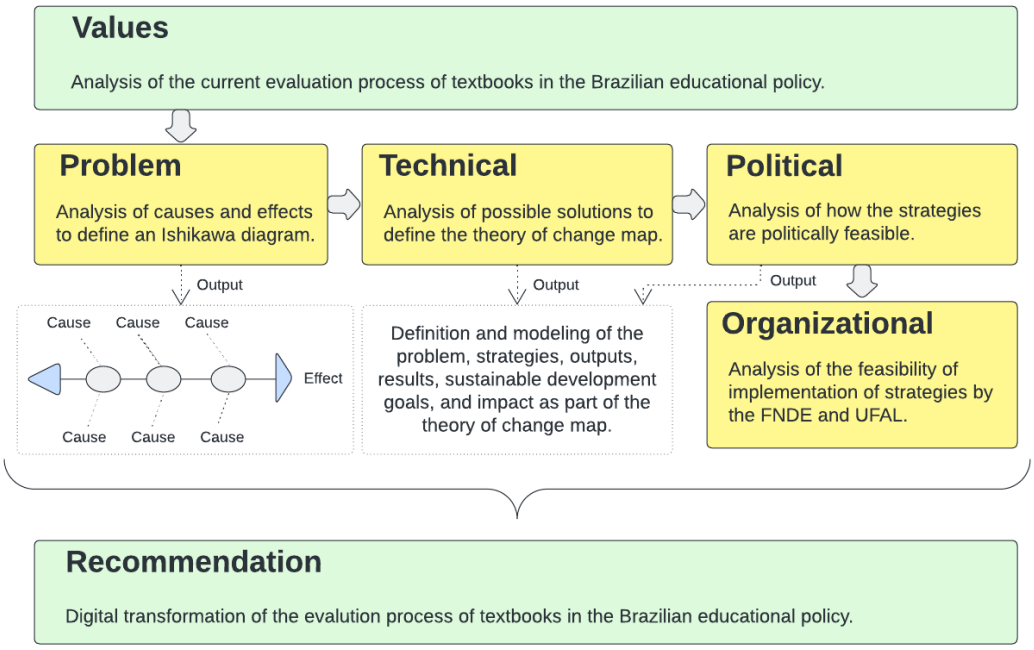

Figure 2. Policy Design Arc-based methodology of the study.

In our study, the theory of change maps all outputs achieved based on the Policy Design Arc framework. The following sections provide a brief background on the Ishikawa diagram, Cohen's Kappa statistic, and the theory of change.

3.1. Ishikawa Diagram

An Ishikawa diagram (also known as a cause–effect diagram or fishbone diagram) depicts the root causes of existing problems in a process or product. This diagram assists in identifying and categorizing the potential causes of the issues by providing a visual representation. The modeling starts by defining the problem to enable backward mapping and grouping causes. Each identified cause is supported by evidence to increase confidence in the analysis. Professor Kaoru Ishikawa developed cause–effect analysis in the 1960s. We interviewed stakeholders and analyzed internal documents (including reports) to generate evidence.

The Ishikawa diagram is considered a relevant approach to support quality management activities due to the possibility of a precise diagnosis of problems. This approach

has been applied in a variety of domains, such as healthcare [15], production [16], and education [17].

3.2. Cohen's Kappa Statistic

Cohen's Kappa statistic is a relevant tool to measure the level of agreement between judges. Jacob Cohen presented Kappa as a new technique in 1960 [18] to measure inter-rater reliability for qualitative (categorical) items. In this study, we apply Kappa to measure the agreement of validators to present a deeper discussion on the current validation and triage process. Therefore, to interpret results, the strength of agreement based on Cohen's Kappa (k) can be classified using different approaches, such as described by Landis and Koch [19] and McHugh [20]. In this article, we used the following classification [19]:

- poor agreement ($k < 0.00$);
- slight agreement (k between 0.00 and 0.20);
- fair agreement (k between 0.21 and 0.40);
- moderate agreement (k between 0.41 and 0.60);
- substantial agreement (k between 0.61 and 0.80);
- almost perfect agreement (k between 0.81 and 1.00).

3.3. Theory of Change

The theory of change comprises a detailed description of the desired change, including a comprehensive explanation of how and why it is supposed to happen, and the related outputs, outcomes, and impacts. The theory requires one to reason about goals and conditions to achieve the goals. Thus, strategies are defined to support the realization of the outcomes. For instance, the theory keeps track of milestones, documents lessons learned, and improves transparency (a relevant issue for public administration). We also used interviews with stakeholders, internal documents (including reports), and the agreement level between validators to create the theory of change in validating and analyzing the editorial quality attribute process from PNLD.

The term theory of change became well-known due to the work of the Aspen Institute and the Roundtable on Community Change [3,21,22]. The theory of change can be visually depicted by explaining the connections between preconditions, long-term outcomes, indicators, and interventions. This approach has also been applied in a variety of domains, such as healthcare [23], agriculture [24], food security [25], and education [26].

4. Results and Discussions

Interviews with stakeholders and the presentation/generation of evidence of problems guide our comprehensive and critical analysis. We planned and executed four sets of semi-structured interviews and one workshop. The interviews and the workshop occurred from March 2022 to December 2022.

1. The first set of interviews relates to the analysis of the current process, problems, and causes of problems.
2. The second set of interviews relates to the analysis of the level of agreement between validators.
3. The third set of interviews relates to understanding the problems regarding the most recent public calls for books that include digital and accessibility features (i.e., accessible publishing).
4. The fourth set of interviews relates to the validation of strategies, expected outputs, expected results, and impacts.
5. Finally, these interviews are presented and discussed with the FNDE to validate and improve the theory of change.

4.1. Understanding the Process for Textbook Validation

We conducted the first set of semi-structured interviews to analyze the current process for textbook validation. Table 1 describes the questions that guided the semi-structured interviews. We interviewed two general coordinators, one triage advisor, and one triage/quality manager. This first set of interviews occurred in March 2022 using Google Meet. However, we also discussed problems, causes, and evidence freely with stakeholders during interviews in addition to the questionnaire. The discussions helped improve our understanding of the internal process.

Table 1. The questions we used to guide the first set of semi-structured interviews.

Number	Question
1	What is your opinion about the problems? Are the specified problems according to reality?
2	What is your opinion regarding the problems? Are the specified problems in line with reality? What are the causes of issues in the screening process? For each identified cause, is there a relationship with actual validations? Were such causes observed in past validations?
3	What are the causes of problems related to resources? Is there a relationship between each specified cause and real-world validations? Were these types of causes observed in past validations?
4	What are the causes of problems related to communication? Is there a relationship between each specified cause and real-world validations? Has this type of cause been observed in past validations?
5	What are the causes of problems related to the environment? Is there a relationship between each specified cause and real-world validations? Was this type of cause observed in past validations?
6	What are the causes of problems related to diligence? Is there a relationship between each specified cause and real-world validations? Have such types of causes been observed in past validations?
7	What are the causes of problems related to documents? Is there a relationship between each specified cause and real-world validations? Were these types of causes observed in past validations?
8	For each cause and effect, have you identified any inconsistencies in terminology?
9	For each specified cause, do you identify any missing causes?
10	Do you recommend reading any specific internal document for the presented evidence?

Figure 3 depicts the organizational diagram used to guide the process of textbook validation and triage in PNLD. The process is led by two general coordinators of triage/quality, two triage advisors, two triage/quality managers, one help desk supervisor, one team supervisor, and a set of validators for each team. However, the number of teams and validators depends on the currently published public call for books, which is a formal document that presents a set of rules required to be followed by publishers of textbooks. Each public call for books is published for a specific object (a type of material) in a call for books.

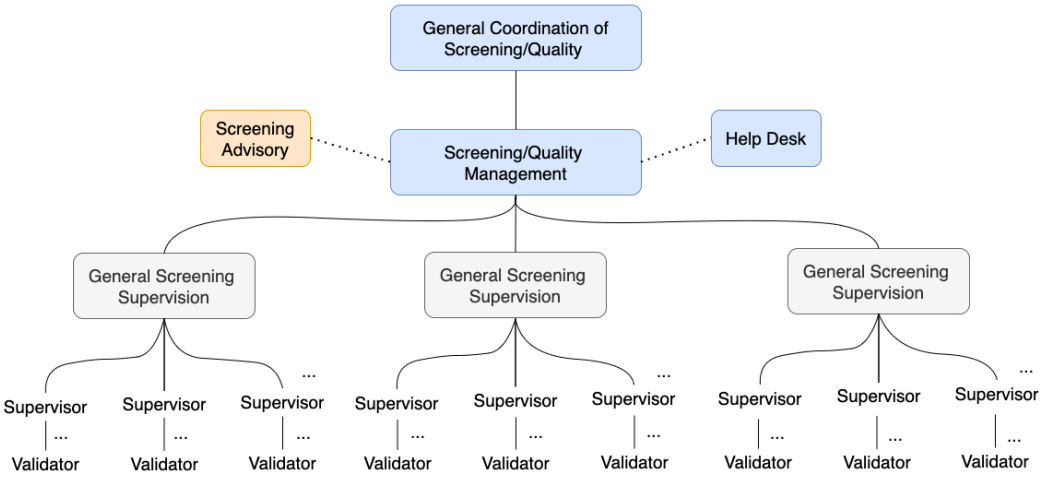

Figure 3. Organizational diagram of the validation and triage of textbooks in the PNLD.

Figure 4 illustrates a flow diagram presenting an overview of the internal process for the validation and triage of books and other digital didactic materials in the PNLD for coordinators and triage supervisors. Four main artifacts are generated using spreadsheets and text editors: a list of textbooks under validation, a list of textbooks distributed to teams, a list of attributes used during evaluations, and supporting materials. The process starts when the general coordinator receives the textbooks from the FNDE. Afterward, the coordinator and triage supervisor define metrics (e.g., number by teams) and distribute the textbooks to each team. The triage supervisor also analyzes the public call for books, prepares a checklist for validations, and creates training materials for validators. They also assign validators to teams and specific supervisors, set up Google Drive folders for each supervisor, and define the validation schedule.

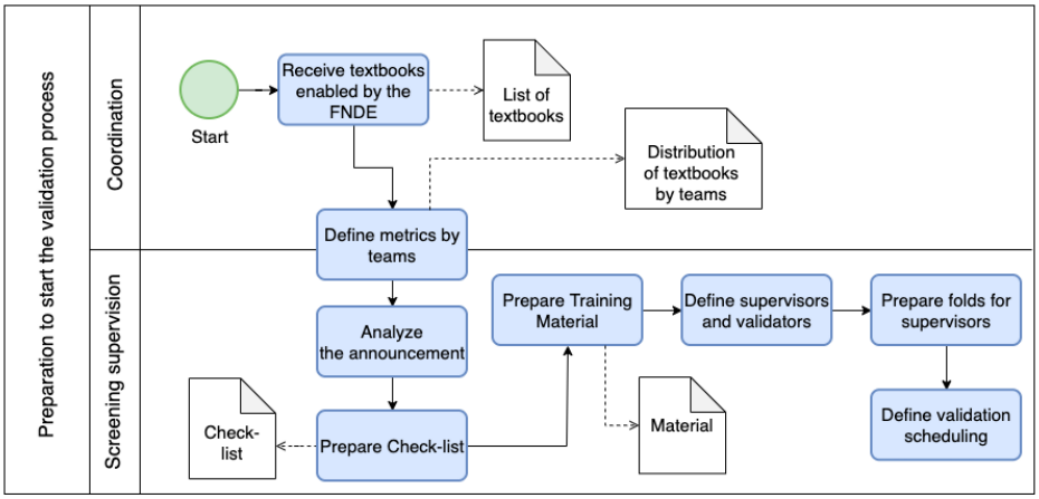

Figure 4. Flow diagram with an overview of the internal process, for coordinators and screening supervisors, of validation and analysis of attributes of textbooks in the PNLD.

Therefore, the first weakness highlighted in this article is the need for updated and relevant process indicators (e.g., number of failures per validator and time spent per textbook per validator). Currently, only a few process indicators are monitored, such as the number of validation failures identified by evaluators during the pedagogical phase, which is an analysis phase that occurs one step ahead. Another example of a process indicator already monitored is the number of failures in textbooks reported by validators (referred to as "diligence").

Once the planning and preparation steps are completed, the validation/attribute analysis begins. Two validators review the textbooks and report any existing failures as diligence. Each team's supervisor consolidates the validations, discarding incorrect failures (false positive diligence). However, due to limited time for consolidation, the team's supervisor only reviews a sample of the validations. During the interviews, we found that disagreements between validators are common. Although screening/quality managers also review the work of the team's supervisors, we confirmed that supervisors might not identify some mistakes at this stage of the process. When failures are consolidated, the publishers are asked to make the necessary corrections. However, some failures persist and are identified by evaluators in the subsequent evaluation phases (e.g., the pedagogical phase).

4.2. Cause–Effect Analysis

We also used the first set of semi-structured interviews (guided by the questions presented in Table 1) to identify and validate problems and causes of problems. The reduced quality of deliveries (a relevant problem) can sometimes occur due to mistakes made by validators that are not identified by supervisors, screening/quality managers, and coordinators. The semi-structured interviews helped in identifying, documenting, and validating all effects (problems) and related causes.

Figure 5 illustrates the reduced time required to complete tasks (another relevant problem). It also provides an overview of high-level process tasks: (1) preparing the resources required to start the process (preparation to start), (2) conducting tasks required before starting double validation (validation starting point), (3) conducting double validations (double validation), (4) inserting validation data into an online validation platform used by publishers and validators, known as PNLD Digital (insertion of data using the platform), (5) requesting and monitoring diligence (diligence), (6) consolidating validation results (consolidation), and (7) reviewing overall results (general review). The most time-consuming task is double validation; however, unexpected situations during the internal process usually delay other tasks.

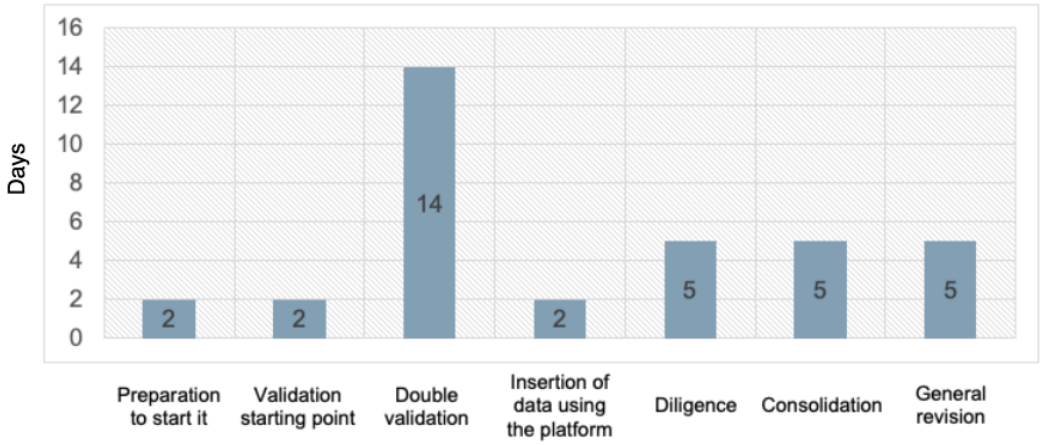

Figure 5. The usual time required to conclude the tasks in the internal process (number of days by task).

4.2.1. Screening

Five causes are directly related to the screening process: inconsistent analysis of textbooks (Cause 1), incorrect understanding of quality requirements from the public call for books (Cause 2), short-term screening carried out for each object (Cause 3), unforeseen business rules throughout validation (Cause 4), and inadequate reports (e.g., with no details) in screening (Cause 5).

Cause 1 of screening exists because incorrect screening is common among validators, as identified by supervisors. Cause 2 of screening compromises the screening process conducted by validators because it depends on the correct understanding of requirements.

Cause 3 harms the screening process due to the large number of textbooks (often hundreds and thousands) required to be validated in a short period. Cause 4 of screening is relevant because validators may conduct validations based on outdated rules when a requirement suddenly changes during the screening process. Cause 5 of screening is prevalent and has a significant harmful impact during the screening process. One recurrent problem is the need for more standardization in reporting textbook failures in a spreadsheet, resulting in a lack of clarity and details in justifications for including the failure.

4.2.2. Resources

Six causes are directly related to the resources used by stakeholders: a high number of manual tasks (Cause 1), lack of consistency in spreadsheets (Cause 2), difficulties in handling spreadsheets, e.g., generating reports (Cause 3), low flexibility when using the platform (Cause 4), recurring problems on the platform due to its recent implementation, e.g., crashes (Cause 5), and lack of understanding by the publisher about the platform's functioning (Cause 6).

Cause 1 of resources is critical because the burden of questioning respondents (e.g., validators responding to checklists) can have effects related to age, level of education, workplace, and type of question [27]. There is evidence of the prevalence of respondent burden due to the length of questionnaires. Additionally, a high number of manual tasks can result in reduced quality in decision making (e.g., precision), reduced performance in decision making (e.g., time), increased efforts during validations, and low-efficiency business processes [28]. In the internal process, many textbooks are validated quickly as manual tasks.

The screening supervisor develops the checklist spreadsheets by reading the public call for book attachments and carrying out a copy-and-paste task (an error-prone task). This task may also be compromised due to a lack of clarity and ambiguous sentences in the public call for books, which can confuse the validators. Some items are adjusted in the checklist spreadsheet to improve clarity.

Causes 2 and 3 of resources relate to known problems of spreadsheets, such as quantity (limited amount of data), linearity (nested and non-linear analyses that are difficult to understand), quality (common calculation errors), presentation (images and graphics are placed inside the spreadsheet), and sharing (after exportation, a spreadsheet may have different versions of components) [29]. Based on the interviews, we identified that, in many situations, manual data collection and report generation requires many hours or days.

For Causes 4, 5, and 6 of resource, it was evident that relevant requirements need to be included in the current platform. The platform still requires too much time for execution and often crashes, necessitating external developers' intervention. Additionally, the platform contains some outdated functionalities and frequently encounters failures. We also identified recurrent communications from publishers with doubts about the platform's functioning. Sometimes, required input fields are not enabled, making it impossible for publishers to correct diligence.

4.2.3. Communication

Three causes are directly related to communication: changes in the requirements of a public call for books (Cause 1), communication between FNDE, UFAL, and other stakeholders (Cause 2), and communication with publishers (Cause 3). For all causes, communication failures negatively influence the project's success, resulting in a lack of clarity [30]. Communication failures are also usually related to the inadequate use of communication technologies. They may involve a macrosocial system level (between various stakeholders at the organizational level) and an organizational system level (failure of shared understanding).

For Cause 2 of communication, we identified at least one situation where there was a lack of a clearly defined communication process between FNDE and UFAL regarding changing requirements in the public call for books, which were only updated online. For

Cause 3 of communication, we identified that communications between UFAL and publishers are conducted through many emails, making it difficult to document. Therefore, the lack of communication becomes a problem that can impact the entire process, reducing the quality of validations. Validators, supervisors, and coordinators should be well-informed about changes in requirements during the process to prevent inconsistent validations.

4.2.4. Environment

Two causes are directly related to the environment: a workplace with distractions (Cause 1) and a home office (Cause 2). Causes 1 and 2 of the environment may harm the performance of validators, supervisors, and coordinators. Examples of impacts include not performing as well as expected, reduced productivity, and limited career improvements. Furthermore, a recent study has shown some problems Brazilian workers face in a home office, including factors such as children at home and internet access [31]. In the validation and triage process of textbooks, validators can work remotely. Although this problem is hard to address and control, it should be considered by Brazilian educational policymakers. Indeed, issues such as children at home may negatively impact the quality of textbook validations.

4.2.5. Diligence

Five causes are directly related to diligence: conflicting validations (Cause 1), consolidation failure (Cause 2), unplanned repeated diligence (Cause 3), inadequate correction of diligence by the publisher (Cause 4), and missed diligence response deadlines by the publisher (Cause 5). Cause 1 of diligence is critical to the quality of the whole process. Group decision making involves many complex and conflicting intrinsic aspects of individuality and human nature [32]. Conflicts are natural when experts make decisions and should be mediated and resolved. During interviews, we identified that conflicting validations by validators are generally resolved by supervisors, given the current time restrictions imposed by the Ministry of Education.

Supervisors should identify such conflicts (Cause 2 of diligence); however, some validation conflicts or incorrect validations are only identified by general screening supervisors, screening/quality managers, or even the Ministry of Education (in the worst case). For Cause 2 of diligence, there is a burden on supervisors who only re-evaluate a sample of the validations to improve quality. This usually happens because of the high number of textbooks and the reduced time for validations. For example, in object 2 (literary books) from the 2022 public call for books, three failures were identified by the Ministry of Education, while in object 1 (didactic books) from the 2023 public call for books (https://bit.ly/3OznYzC, accessed on 27 May 2023), six failures were identified by the Ministry of Education in the subsequent step of pedagogical evaluation. Failures are very concerning because if the Ministry of Education staff cannot identify incorrect validations, some failures may remain, and the students and teachers may receive textbooks with quality problems.

Causes 3 and 4 of diligence usually occur when the publisher fails to provide the required data in the platform, requiring more time to complete the corrections. For example, if we consider an estimate, approximately 50% of the problems related to time are a result of missing fields when publishers input data into the platform. For Cause 5 of diligence, considering the report of diligence related to object 3 in the 2023 public call for books, more than 3.37% of the collections of textbooks were delayed, with the validation conducted in the first semester of 2022.

4.2.6. Document

Four causes are directly related to documents: lack of clarity of the public call for books (Cause 1), lack of completeness of the public call for books (Cause 2), inconsistency between the checklist and the public call for books (Cause 3), and reduced clarity in checklist items (Cause 4). For Causes 1 and 2 of the documents, during interviews, we identified that some

requirements of the public call for books appear to be outside the scope of the validation and analysis of editorial quality attributes.

For Causes 3 and 4 related to the document, we identified that inconsistency in the checklist and reduced clarity in checklist items are common problems during the internal process and may result in delays of up to 8 days. Given the limited time available for validations, such delays cannot be acceptable. As an example of Causes 3 and 4 related to the document, there is no exclusion code for digital materials on the platform in validating the literary works from the 2023 public call for books. The lack of exclusion codes complicates the validation and triage process and the consistency of documentation.

To summarize our findings, Figure 6 presents the Ishikawa diagram for validating and analyzing editorial quality attributes of didactic books, literary books, and other digital didactic materials. We developed and validated such a diagram based on interviews with stakeholders. The risks of reduced quality in deliveries with increased time to conclude tasks are critical problems that can negatively influence the results of the process realization.

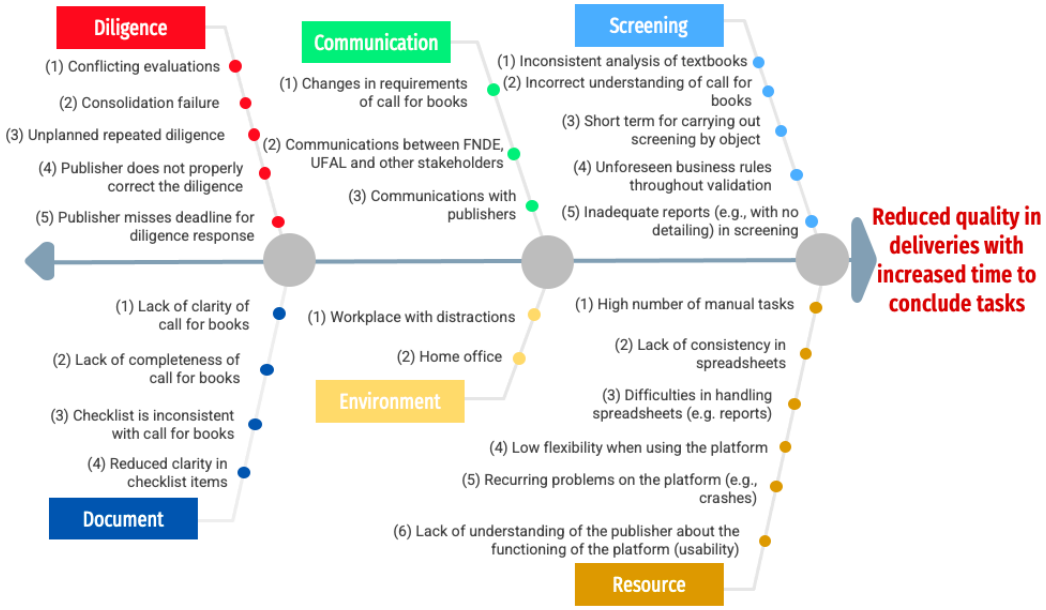

Figure 6. Causes and effects, as an Ishikawa diagram, of the validation and analysis of attributes of textbooks in the PNLD.

4.3. Agreement Analysis between Validators

We used Cohen's Kappa statistic to compute the agreement level for each diligence between validators of textbook attributes to support the comprehensive analysis of the validation and triage process. We used the SPSS statistical software to conduct the analyses.

Figure 7 illustrates the results of Cohen's Kappa for each of the nine teams that evaluated textbooks from the PNLD 2023 public call for books, object 03. Two different validators (i.e., two rounds) evaluated the same textbook to improve confidence in validations for each team. The highest agreement was achieved by teams 3, 4, and 7, who showed a fair level of agreement with validators while validating textbooks. However, the negative results for other teams are alarming and indicate flaws in the validation and triage process. It does not necessarily mean that all textbooks were not correctly corrected at the end of the process, but it suggests that more time is required to complete the process, and some textbooks may not be corrected as expected.

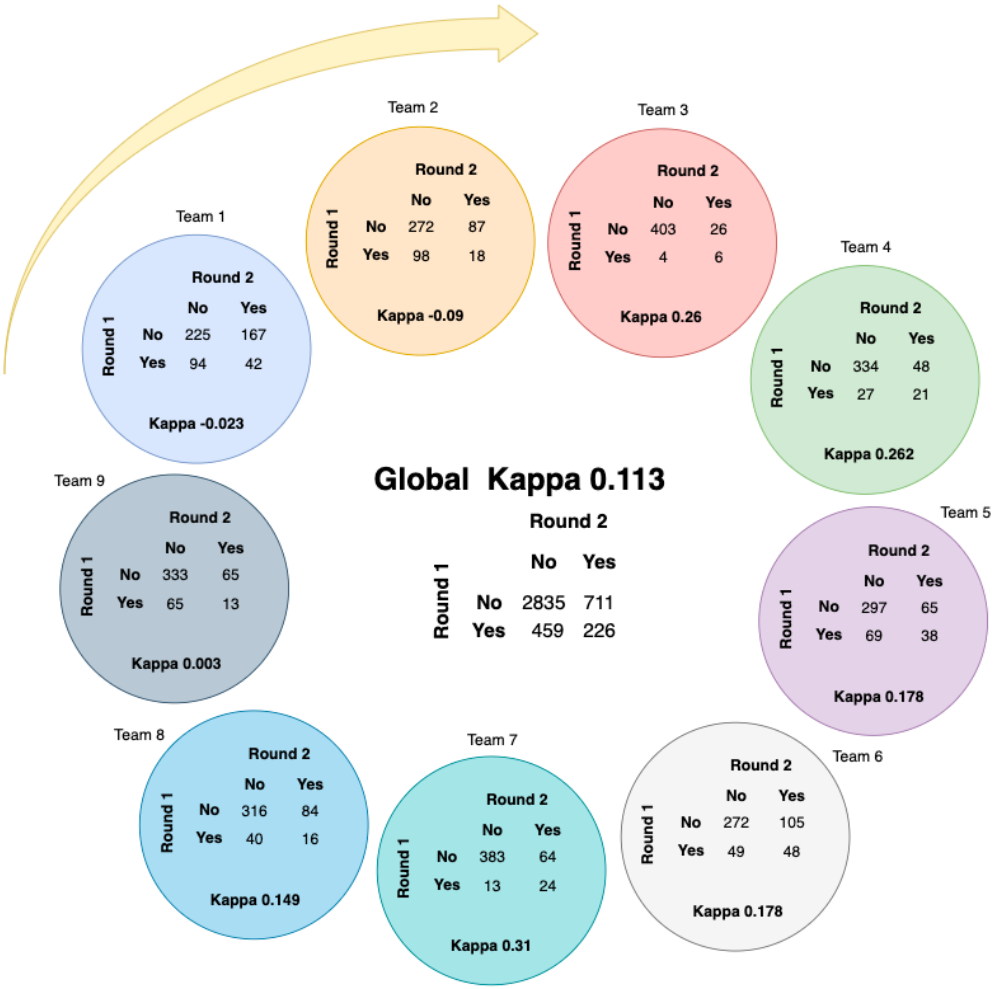

Figure 7. Cohen's Kappa results by teams from the PNLD 2023 public call for books, object 03.

Given the low Kappa results, we conducted a second set of interviews to understand the problem. Table 2 describes the questions used to guide the second set of semi-structured interviews. However, during the analysis, we observed that sometimes validators recorded failures in the checklist but did not include them in their final report (analyzed by supervisors). Such missing data, probably caused by input mistakes due to a lack of attention, may negatively impact the quality of textbook validations.

Table 2. The questions we used to guide the second set of semi-structured interviews.

Number	Question
1	Are the Kappa results low due to a training problem?
2	Could the low Kappa results be attributed to the lack of commitment from validators?
3	Are the Kappa results low due to the reduced time to conclude the tasks?
4	Are the Kappa results low due to the high number of textbooks to evaluate?
5	Are the low Kappa results due to inconsistency or ambiguity in the public call for books, such as unclear exclusion criteria?

Based on interviews with the screening/quality management team, we identified that the main problems negatively impacting the level of concordance are the lack of training, the commitment of validators, and unforeseen business rules throughout the validation

and triage process. For example, recruiting new (inexperienced) validators is related to the need for more training. Another problem related to training is the time validators spend without performing project tasks before being called to work as validators in a specific public call for books. Additionally, commitment is a recurring problem among validators.

During the interviews, the time to conduct the validations did not appear to be a possible problem directly affecting the level of concordance. As part of the current process, time estimation is conducted for validations to define the number of validators needed for each team, considering the number of textbooks. Therefore, if there is an impact, it is not the main reason for the low agreement. Based on such estimates, overload is not considered a problem that directly impacts the agreement level.

However, inconsistencies/ambiguities in the public call for books (and consequently, in the exclusion codes) may significantly impact the low levels of agreement. The interviewed staff reported that discrepancies between validations are identified due to unforeseen business rules throughout the validation and triage process. Thus, the disagreements may also result from a communication problem related to changes and misunderstandings of requirements.

We also analyzed the validation results between validators after the supervisors' and screening/quality managers' consolidations. At this stage of the process, the agreements remained low for all teams: team 1 (0.431), team 2 (0.0289), team 3 (0.264), team 4 (0.38), team 5 (0.319), team 6 (0.373), team 7 (0.496), team 8 (0.148), and team 9 (0.18). The consolidation stage increased the global Kappa from 0.113 (slight agreement) to 0.336 (fair agreement). The global Kappa increased due to the inconsistent validations identified by supervisors and screening/quality managers. Such inconsistent validations were removed (as not reasonable diligence) or integrated (as the same diligence) during the supervisors' and screening/quality managers' consolidations. For instance, in some cases, validators reported the same failure differently, subsequently integrated by supervisors and screening/quality managers as unique diligence.

Therefore, to improve the discussion and further understand the previous concordance results, we analyzed the validation results for each validator by comparing their decisions with the final consolidation of supervisors and screening/quality managers. We used Cohen's Kappa statistic to analyze the performances of each validator from the nine teams that evaluated textbooks based on the PNLD 2023 public call for books, object 03. Each team comprised nine or ten validators in such a public call for books. It is relevant to note that the number of validators may vary for the validated object. Only a few validators presented an almost perfect or substantial agreement with the consolidation results, and some of the validators needed to show better agreement with the consolidation results. Such results strongly evidence some of the causes described in the Ishikawa diagram (i.e., Figure 6), such as inconsistent analysis of textbooks. For instance, Figure 8 illustrates Cohen's Kappa results by comparing each validator of team 1 and the consolidation of supervisors and screening/quality managers. Validators 3 and 6 were the most reliable, agreeing substantially with the consolidation results. In contrast, validator 7 achieved a moderate agreement, while validators 5, 8, and 9 achieved fair agreement, and validators 1, 2, and 4 achieved slight agreement.

However, it is not possible to conclude whether low commitment explains such results. We argue that validators' low performance results from problems such as low commitment and incorrect understanding of the public call for books. These findings indicate that proposing strategies is necessary to reduce the negative impacts of this type of real-world problem. The proposed strategies should consider solving automation, process, training, and human resources problems.

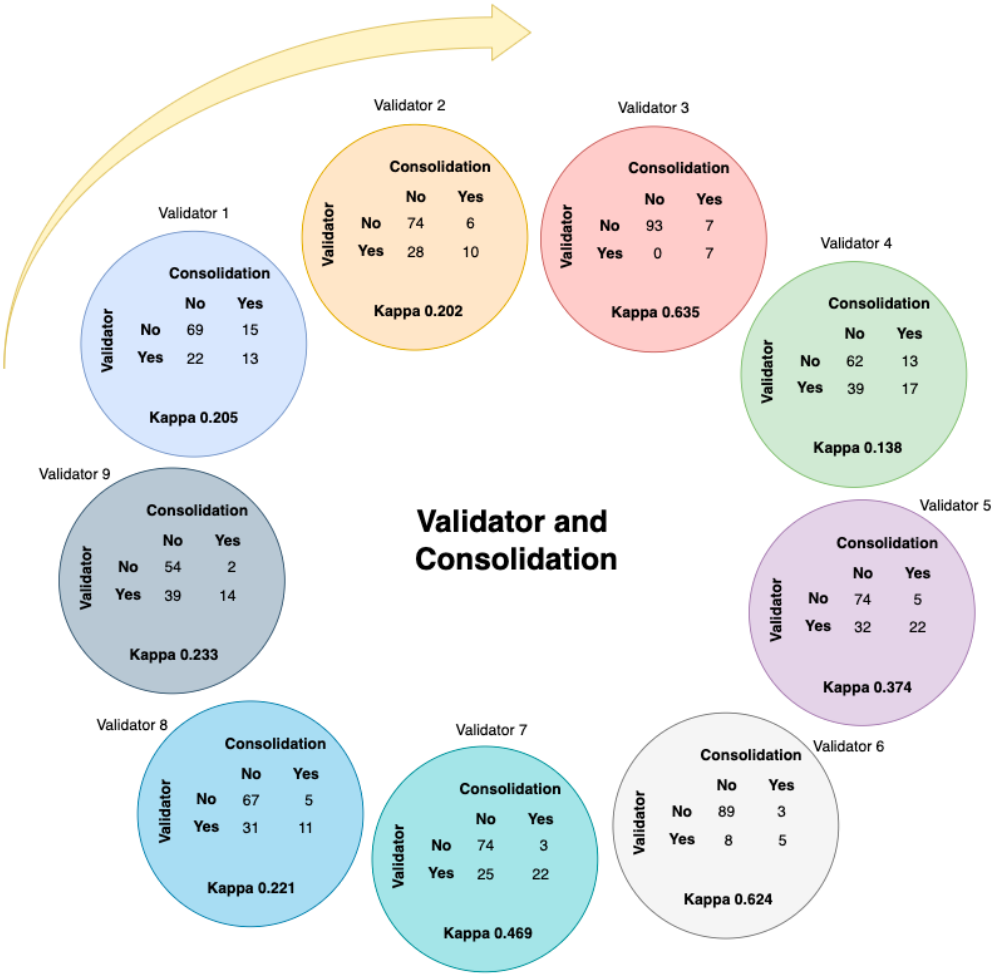

Figure 8. Cohen's Kappa results by comparing each validator of team 1 and the consolidation of supervisors and screening/quality managers. Such results relate to the PNLD 2023 public call for books, object 03.

4.4. Digital Books and Accessibility

We also analyzed the recent public calls for books, namely PNLD 2023, Object 01; PNLD 2023, Object 01 Accessibility (AC); PNLD 2024, Object 01; and PNLD 2024, Object 03. Requiring digital books and accessibility features was a recent and significant initiative of the PNLD. We aimed to understand the impact of validating digital books and books with accessibility features, which can support improvements in future public calls for books. We read the public calls, interviewed participants in the triage process, and applied statistical methods based on data recorded during the triage process. The availability of high-quality digital books is crucial to increase equity in Brazilian public schools, helping achieve relevant SDGs.

Figure 9 presents the global Kappa results for all the public calls for books analyzed in this article, including physical, digital, and accessible books. We observe that the level of agreement was higher for the analysis of HTML5 documents (i.e., digital books) after the inclusion of automatic validations (by software) of HTML5 elements required by the PNLD 2024 OBJ3 VI public call for books, indicating the possible benefit of using decision support tools as part of our theory of change presented in Section 4.5. However, for accessibility, the level of agreement was poor, indicating that many problems occurred during the analysis of the books. It is important to point out that the time available for validating books in the PNLD 2023 public call for books was shorter than the time normally available for the execution of the process due to internal reasons. This negatively impacted the execution of the process, consequently reducing the agreement results.

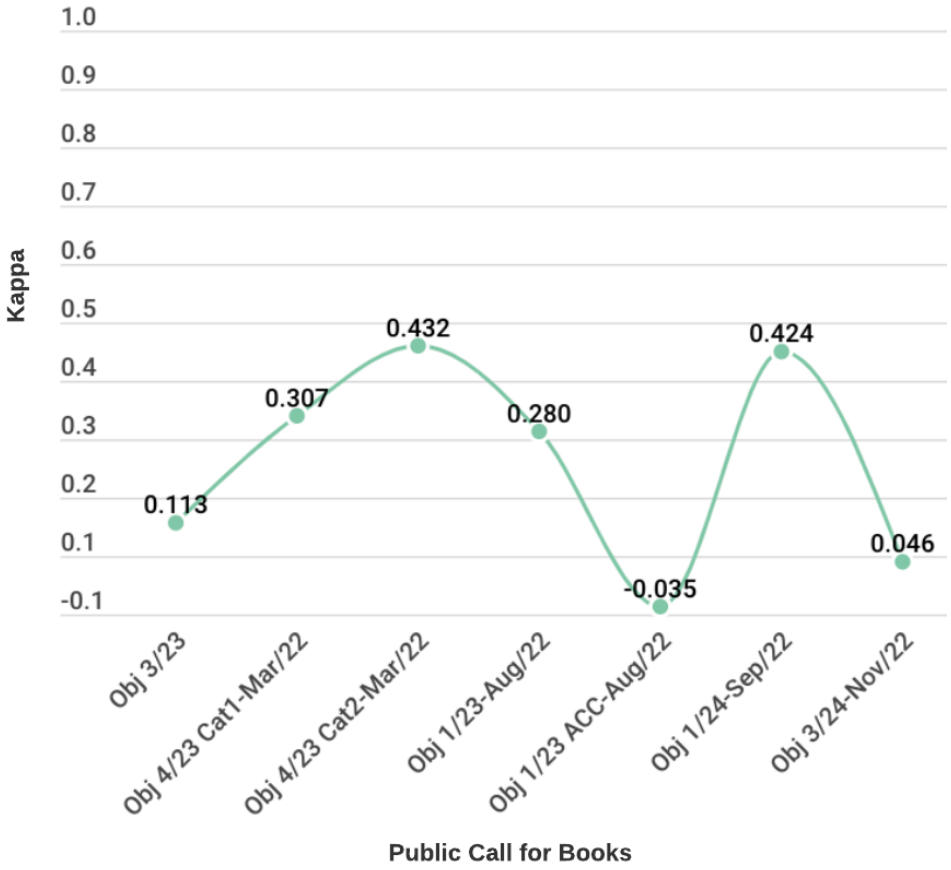

Figure 9. Global Cohen's Kappa results for the public call for books analyzed in this article.

The agreement results for physical books were higher than those for digital books. One hypothesis that may justify this difference is the level of experience of validators in previous public calls for books (i.e., physical books) that were conducted before requiring digital versions of books in the most recent public calls. Thus, including digital books and accessibility features resulted in new challenges during the triage process. These challenges, in part, are related to the level of detail and clarity of the public calls for books, especially in descriptions of exclusion codes.

During our analysis, we found that the PNLD 2024 public call for books showed improvements compared to the PNLD 2023 public call for books. However, several aspects can still be improved to increase clarity in item descriptions and exclusion codes, thus reducing possible problems during the process. For instance, the description of exclusion codes can impact both validators' and editors' understanding of the requirements.

To gain a deeper understanding of the potential impact of digital books and accessibility in public calls for books, we conducted semi-structured interviews with six supervisors (including those responsible for accessibility), three validators (including those responsible for accessibility), one coordinator, and one accessibility manager. Understanding the perspectives of stakeholders actively involved in the triage process is important for identifying issues and recommending improvements based on a more comprehensive understanding of the current process.

We defined the questions presented in Table 3 based on the preliminary analysis of the most recent public calls for books of the PNLD. Our aim was to address problematic topics and encourage the interviewed actors to present possible causes of problems and solutions. Therefore, it is possible to observe that the completeness and clarity of the requirements, such as exclusion codes, described in the public calls for books are often highlighted as critical problems. The descriptions of exclusion codes should be carefully revised to increase the level of clarity and consequently reduce errors, both by editors

and validators. It is recommended to revise accessibility-related requirements carefully to prevent clarity and completeness problems in future public calls for books.

Table 3. The questions we used to guide the third set of semi-structured interviews.

Number	Question
1	What is your perception of the quality of the exclusion code descriptions included in the most recent public call for books?
2	Do you believe that all the necessary exclusion codes have been included in the most recent public call for books? Or do you think that there are some codes that are missing?
3	What is your perception of the shift from validating only physical books to digital books? Were any specific difficulties identified?
4	What is your perception of including accessibility in the most recent public calls for books? Were any specific difficulties identified?
5	What are the main difficulties identified during the current triage processes?
6	What can be improved/adjusted during the current triage processes?

In addition to conducting interviews, we analyzed the number of failures identified in books during the triage processes, considering the exclusion code. This type of analysis aimed to identify the codes presenting the highest occurrence rate and recommend improvements to provide clear and more detailed public calls for books. Figure 10 shows the number of failures by exclusion code related to the PNLD 2023 public call for books, Object 1. The most frequent fault codes were:

- AFDid27: The book presented (offered) a disagreement with the editorial structure provided by the FNDE.
- AFDid26: The digital book presented (offered) a disagreement with the public call for books.
- AFDid17: The book specifications are not equal to those provided in the FNDE platform.

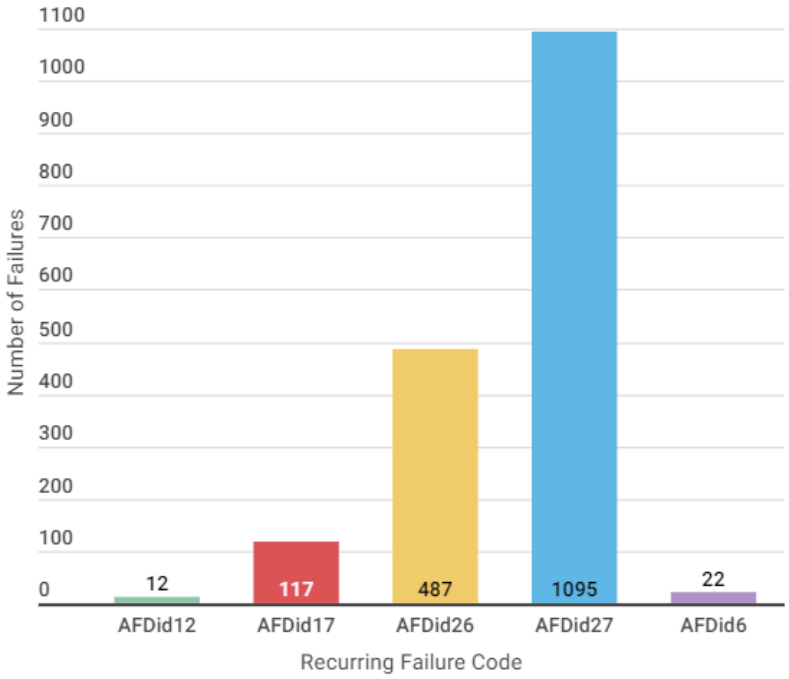

Figure 10. Most recurring failures by exclusion code (PNLD 2023 Object 1, physical attributes).

Figure 11 shows the number of failures by exclusion code related to the PNLD 2023 public call for books, Object 1 AC. The most frequent fault codes were:

- AFDidAC10: There is a lack of descriptions of images and complex objects (such as formulas, infographics, and diagrams).
- AFDidAC13: Some tables are not accessible.
- AFDidA14: The colors used for highlighting text or frames (boxes) are at odds with the level of contrast required for reading by users with impaired vision.

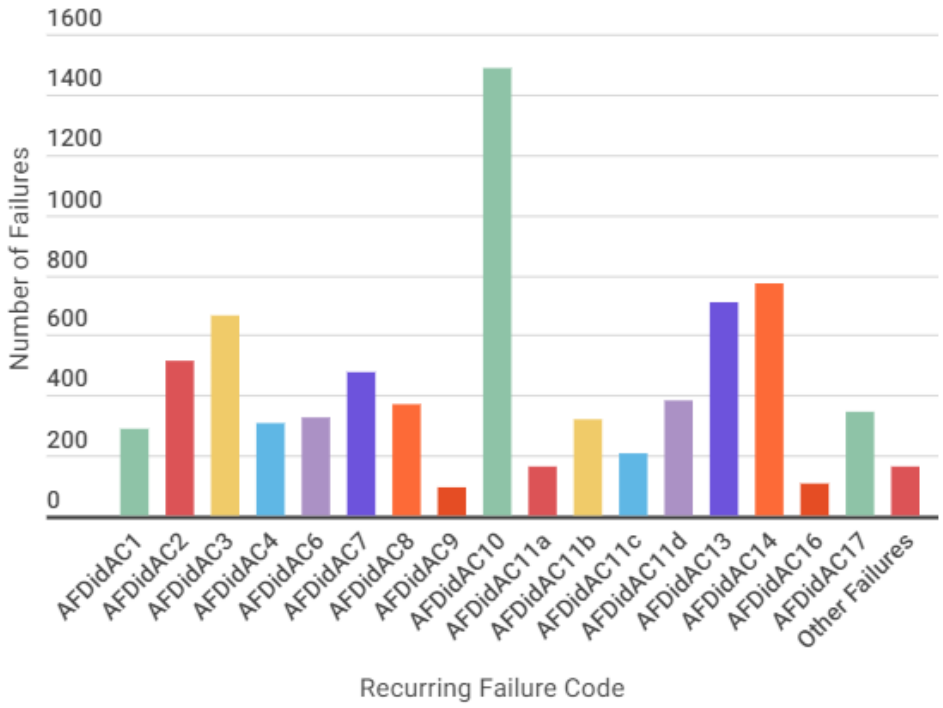

Figure 11. Most recurring failures by exclusion code (PNLD 2023 Object 1 AC).

The codes for other faults identified are described below:

- AFDidAC1: The book disagrees with the accessibility guidelines specified by FNDE.
- AFDidAC2: The lists do not present a logical order in the book.
- AFDidAC3: The text's heading levels are not in logical order.
- AFDidAC4: The sections do not have a logical order in the book.
- AFDidAC6: Page numbering structure does not follow the public call for books specifications.
- AFDidAC7: The links are not accessible.
- AFDidAC8: Glossary disagrees with the public call for books specifications.
- AFDidAC9: Footnotes are not properly linked in both directions.
- AFDidAC11a: There are contents/elements not accessible in the body of the book: (a) position of the elements.
- AFDidAC11b: There are contents/elements not accessible in the body of the book: (b) word search exercises.
- AFDidAC11c: There are inaccessible contents/elements in the body of the book: (c) indications incompatible with the digital book (e.g., "do not write in this book").
- AFDidAC11d: There are inaccessible contents/elements in the body of the book: (d) words/phrases misread by NVDA.
- AFDidAC16: Landmarks are missing.
- AFDidAC17: Other issues found.

Figure 12 shows the number of failures by exclusion code related to the PNLD 2024 public call for books, Object 1 AC. The most frequent fault codes were:

- VDidC23: The book specifications differ from those informed on the FNDE platform.
- VDidC33: The digital book does not meet the required size/duration specified in the public call for books.
- VDidC34:The digital book does not include usage guidelines for its contents as required by the public call for books.
- VDidC35: The number of elements in the digital book does not match the specifications provided in the public call for books.
- VDidC4: The file uploaded to the platform does not match the item filled out.
- VDidD23: The title page contains text or illustration.
- VDidD24: The unidentified digital book needs to comply with the required size/duration specified by the public call for books.
- VDidD25: The unidentified digital book does not provide content usage guidelines.
- VDidD26: An unidentified digital book has a different number of elements than required by the public call for books.
- VDidD5: The book in PDF format does not comply with the specifications presented in Annex II of the public call for books.

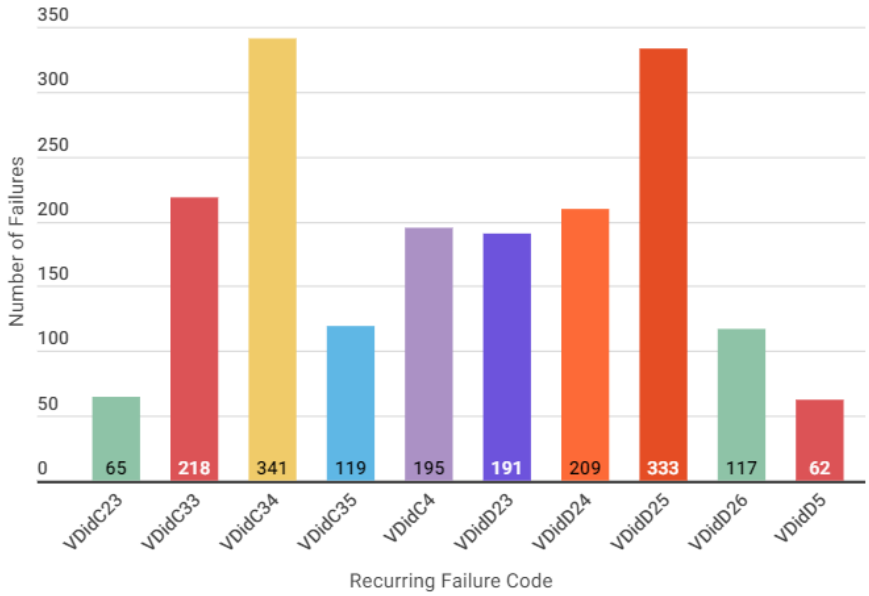

Figure 12. Most recurring failures by exclusion code (PNLD 2024 Object 1 VI).

Finally, Figure 13 presents the number of failures by exclusion code related to the PNLD 2024 public call for books, Object 3 VI. The fault code with the highest occurrence was VLit32 (the number of elements in the digital book is not following the public call for books). The codes for the other identified faults were:

- VLit5: The PDF version of the book does not conform to the specifications outlined in Annex II of the public call for books.
- VLit10: Missing pages.
- VLit11: Pages are swapped, flipped, or out of sequence.
- VLit13: The book does not have proper pagination in its printing.
- VLit15: The book presents blank page(s).
- VLit19: The PDF version of the book disagrees with the book's title, volume title, edition year, or edition number.
- VLit20: The book was submitted without the catalog sheet, the front cover image, or the documentation required in Annex VIII of the public call for books.
- VLit21: There are mandatory blank fields on the PNLD platform.
- VLit22: The book specifications differ from those informed on the PNLD platform.

- VLit23: The authorship(s), year, edition number, and publisher data in the catalog file do not match those registered on the PNLD platform.
- VLit25: The year, edition number, and publisher listed in the catalog file do not match those in the description of the submitted book.
- VLit29a: The back cover does not contain the ISBN.
- VLit29b: (b) The ISBN is superimposed on a colored background on the cover.
- VLit33: The digital book has flaws in navigation.
- VLit34: The digital book does not reproduce the entire content of the printed book.
- VLit37a: The teacher's digital book support material does not conform to the guidelines outlined in the public call for books; specifically, the support material exceeds the page limit of 15 to 30 pages.
- VLit37d: (d) Supporting material does not contain the context of the book.
- VLit37e: (e) Supporting material does not present the artistic nature of the book.

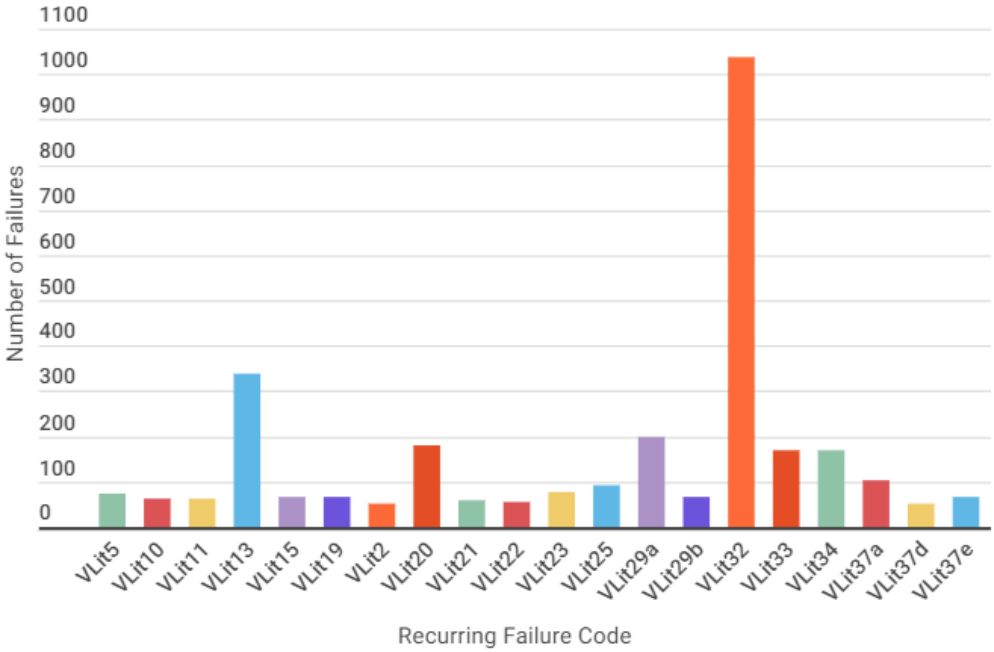

Figure 13. Most recurring failures by exclusion code (PNLD 2024 Object 3 VI).

Therefore, we recommend special attention to exclusion codes (and code descriptions) with higher frequencies. Publishers can use this information to check for specific failures to mitigate frequently occurring issues carefully. Furthermore, developers of public calls for books (i.e., policymakers) can use this information to improve failure code descriptions to increase clarity, also aiming to mitigate frequent problems. After adjusting the descriptions of the exclusion criteria, it is also necessary to verify if the publishers clearly understand them. This can improve the quality of submissions and speed up the validation and triage process.

4.5. Strategies, Expected Outputs, Expected Results, and Impacts

We used the results from the Ishikawa diagram (causes and effects) to define a theory of change for validating and analyzing attributes of textbooks in the PNLD. Thus, from the problem and causes, we described strategies, outputs, immediate results, intermediate results, main results, and impacts. Firstly, we categorized the causes based on four groups (Figure 14): cannot be addressed, automation, training and human resources, and process.

Therefore, to reason about the digital transformation of the validation and triage process of textbooks in Brazilian educational policy, we only addressed the causes related to automation, training and human resources, and process. We also conducted semi-structured interviews with stakeholders to identify and validate strategies, expected outputs, results,

and impacts (fourth set of interviews). Table 4 describes some questions to guide the fourth set of semi-structured interviews. We interviewed general project coordinators, triage/quality managers, and FNDE staff.

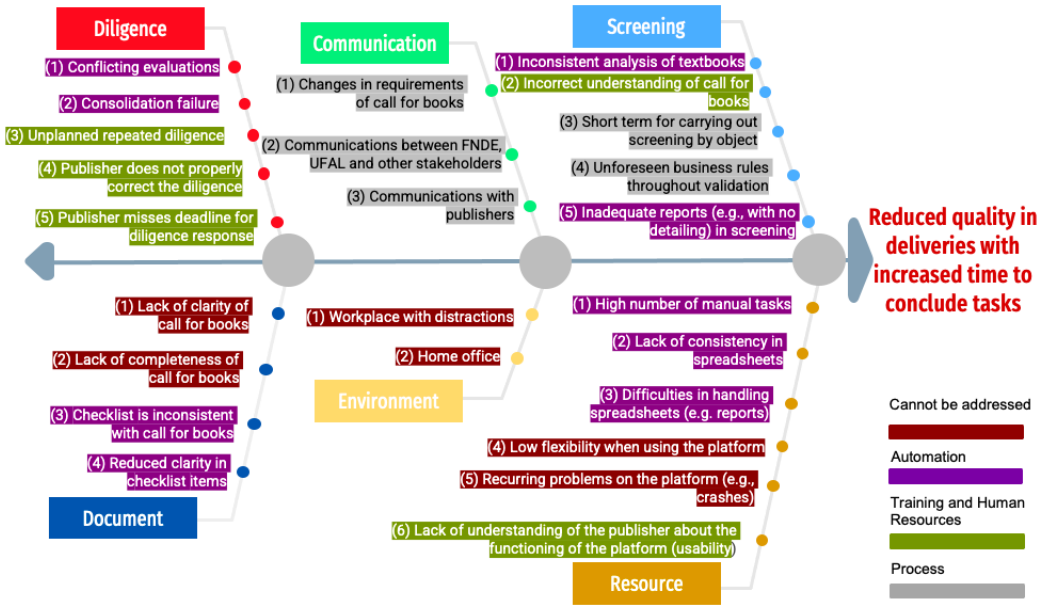

Figure 14. Categorization of causes and effects of the Ishikawa diagram for the validation and analysis of attributes of textbooks in the PNLD.

Table 4. The questions we used to guide the fourth set of semi-structured interviews.

Type	Question
Technical	Do you believe that the proposed strategies follow the needs of the PNLD?
	Do you believe stakeholders (e.g., validators and supervisors) will accept the proposed strategies well?
	Do you believe the proposed strategies may be difficult to implement in the current process (e.g., during an ongoing validation and triage process)?
	Do you have any strategy or modification suggestions for the proposed strategies?
Political	Do you believe the proposed strategies can assist the Brazilian government in achieving sustainable development goals?
	Do you believe the proposed theory of change can positively impact Brazilian educational policies?
	Do you believe the strategies, outputs, and expected results follow Brazilian public policies?
Organizational	Do you believe that the proposed strategies will be well-accepted by the FNDE?
	Do you believe that digital transformation can positively impact the process of validation and analysis of attributes of textbooks?
	Do you believe the suggested indicators will be relevant for the quality management of process tasks?

Figure 15 presents a layered view of the theory of change map for validating and analyzing attributes of textbooks in the PNLD. The bottom layer of the map corresponds to the effects (highlighted in red on the left side of Figure 14) identified in the Ishikawa diagram, which include reduced quality and time as the problem to be addressed. Each block of the layer above (causes) describes the causes related to a specific group, such as automation being the focus of the first block.

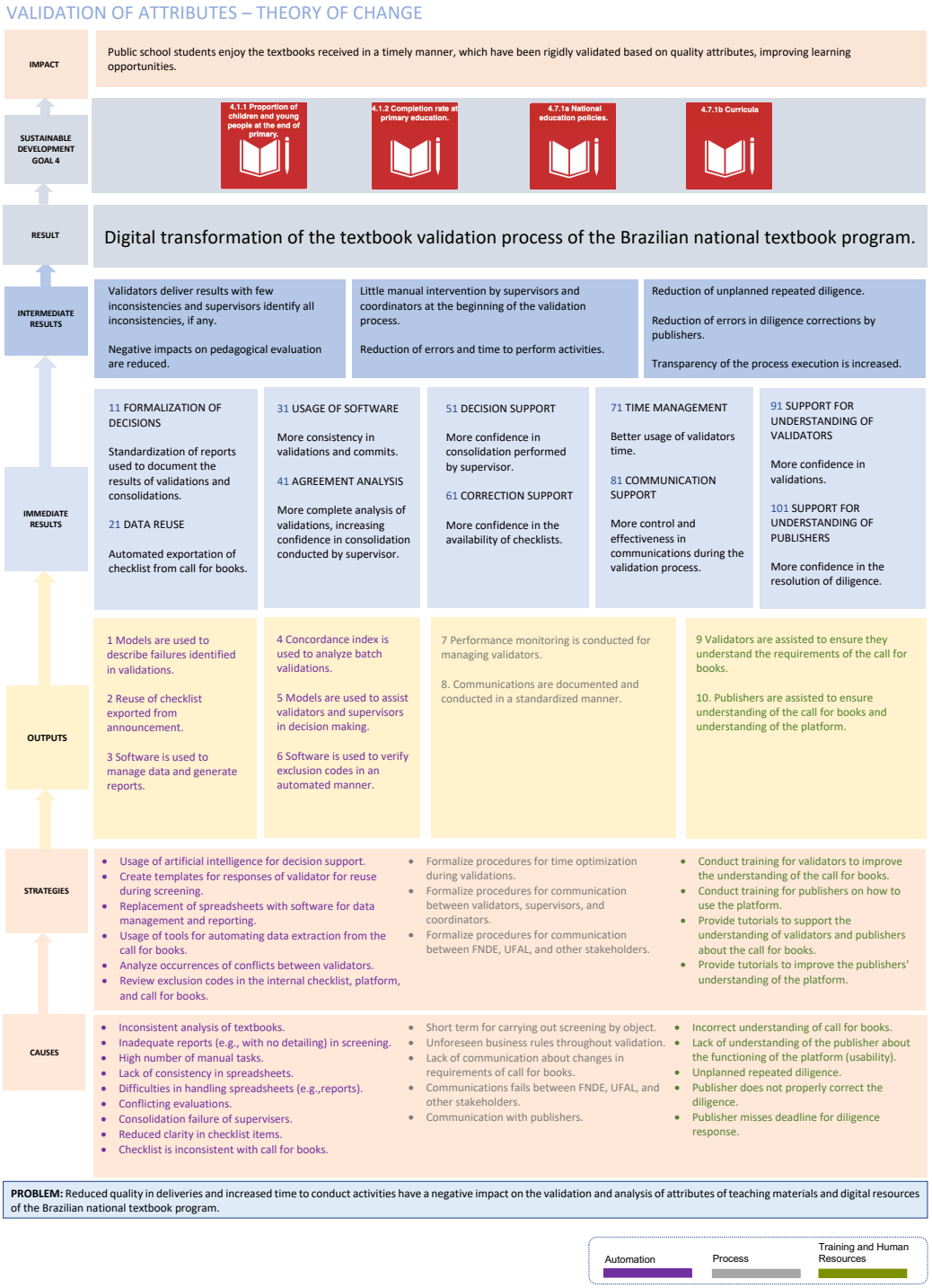

Figure 15. The theory of change to improve the validation and triage process in the PNLD.

4.5.1. Strategies

We analyzed the validation and triage process and defined strategies to address the problems' causes. Augmented intelligence is a relevant strategy to address automation-related causes. The concept of augmented intelligence envisions designing systems that merge humans and artificial intelligence [33]. Augmented intelligence can enable the development of decision support systems to address inconsistent textbook analysis, many manual tasks, conflicting validations, supervisor consolidation failures, and checklist inconsistencies with public calls for books. For example, natural language processing and machine learning can support identifying textbook exclusion criteria problems (e.g., ambiguities) described in public calls for books. Another example is using natural language processing and machine learning to assist supervisors in identifying problems during validations. It is also possible to assist validators in identifying failures such as missing pages and duplicated images through computer vision techniques.

In addition to augmented intelligence, other strategies can improve the quality of deliveries and decrease the time. For instance, the replacement of spreadsheets with software for data management and reporting, the definition of templates for validators' responses, the usage of tools to automate data extraction from a public call for books, and the analysis of conflicts among validators. For example, Cohen's Kappa statistics can be a helpful tool to verify the level of agreement among validators, supporting the consolidation of supervisors.

For process and training and human resources, the strategies formalize procedures for communication and training by the availability of supporting materials/courses, respectively. In this article, our primary focus is discussing the automation group to support the digital transformation of the textbook validation and triage process of the PNLD.

4.5.2. Outputs

Some relevant outputs are enabled by implementing the automation strategies, such as models, automatically generated checklists, software for data management, reports generated by software, concordance index, and automatically verified exclusion codes from exclusion criteria. Additionally, the process strategies ensure that time spent by validators is managed correctly by supervisors and that communications are adequately documented. The training and human resources category is relevant to provide an understanding of a public call for books by validators and publishers. It also supports publishers in understanding the platform.

4.5.3. Immediate, Intermediate, and Main Results

Therefore, based on the digital transformation, we argue that it is possible to achieve immediate results:

1. The formalization of decisions for both validators and supervisors.
2. The reuse of data automatically extracted from a public call for books.
3. The use of software enables more consistent validations and consolidations.
4. The agreement analysis aims to improve confidence in consolidations.
5. The decision support also improves consolidations.
6. The support for correction improves confidence in checklists.
7. Time management to enhance the performance of validators.
8. The support for improving the control and effectiveness of communications.
9. The support to enhance the understanding of the public call for books.

Such immediate results enable expected intermediate outcomes, such as validators delivering validations with fewer inconsistencies, supervisors identifying all discrepancies, if any, and little manual intervention by supervisors and coordinators during the preparation for the validation and triage process (Figure 4).

Other relevant intermediate results include:

- reducing errors and time to conduct tasks for validators and supervisors;

- reducing unplanned repeated diligence;
- reducing publishers' mistakes during diligence corrections.

We consider such intermediate results essential to achieve the digital transformation of the textbook validation and triage process of the PNLD, which is the main expected result.

4.5.4. Impact

The process improvements proposed in this article are expected to benefit Brazilian public school students by providing high-quality textbooks and improving educational equity through the PNLD. Such improvements can positively impact the quality of the teaching–learning process. Therefore, our expected impact can be related to the SDGs report 2021, specifically targets 4 and 9 of the United Nations report 2021 [34]. However, for the purpose of illustration only, the theory of change map in Figure 15 describes Sustainable Development Goal 4, Target 4.1.

4.5.5. Indicators

In Figure 15, we presented two general indicators of the SDGs, target 4.1. Indicator 4.1.1 relates to the proportion of children and young people (a) in the 2nd/3rd year; (b) at the end of primary; and (c) at the end of lower secondary school who must achieve at least a minimum level of proficiency in (i) reading and (ii) mathematics by sex. Indicator 4.1.2 relates to the completion rate (primary education, lower secondary education, upper secondary education).

However, in this study, we also recommend performance indicators to manage the quality of the validation and triage process of textbooks:

- Performance indicator 1: the number of ambiguities in the exclusion codes of the public call for books is equal to or close to zero (e.g., Figure 16a).
- Performance indicator 2: concordance index between validators shall be almost perfect (e.g., Figure 16b). Performance indicator 2.1: number of false positive results.
- Performance indicator 3: the number of failures identified by coordinators is equal to or close to zero.
- Performance indicator 4: the number of requests by supervisors for understanding the decision making of validators is equal to or close to zero.
- Performance indicator 5: the number of failures identified by the Ministry of Education staff is equal to or close to zero. Performance indicator 5.1: the number of failures identified by UFAL is equal to or close to zero.
- Performance indicator 6: the time for delivery of results does not exceed the deadline defined by the FNDE.
- Performance indicator 7: publisher finished the training successfully.
- Performance indicator 8: validator finished the training successfully.
- Performance indicator 9: number of textbooks rejected. Performance indicator 9.1: number of failures not fixed by publishers is equal to or close to zero. Performance indicator 9.2: number of textbooks by the publisher directly rejected. Performance indicator 9.3: number of textbooks by publisher rejected after diligence.

Figure 17 shows an overview of the strategies and proposed performance indicators for the automation category from our theory of change, guided by augmented intelligence techniques (e.g., natural language processing and machine learning).

Based on the interviews, we argue that improving textbook quality can have a positive impact on education quality. However, it is essential to ensure that such textbooks reach public schools. For example, the choice of the textbooks list can also affect target 4. Therefore, an improved and faster validation and triage process can result in delivering books at the right time for students and in providing more textbook options (like a menu of textbooks) for teachers in public schools in Brazil.

Furthermore, the proposed strategies follow the current Brazilian public policies. However, some deployment barriers may be due to validators' and supervisors' acceptance

of automation solutions. Therefore, it is essential to provide information about the purposes of automation solutions. More interactions are also necessary between deployers and those who conduct validations and supervisions. Further studies can be conducted during deployments to understand and mitigate acceptance problems.

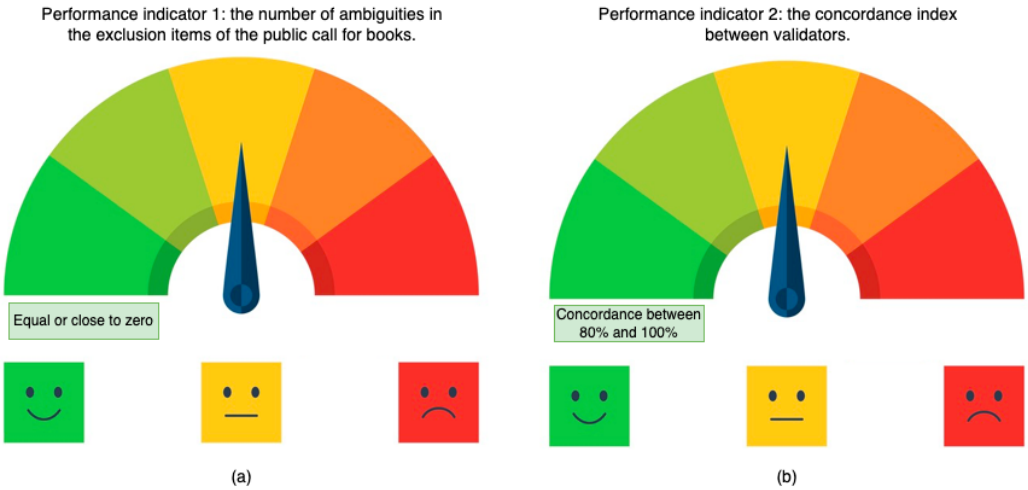

Figure 16. Illustrations of performance indicator 1 and performance indicator 2. (**a**) The performance indicator related to the number of ambiguities. (**b**) The performance indicator related to the concordance index.

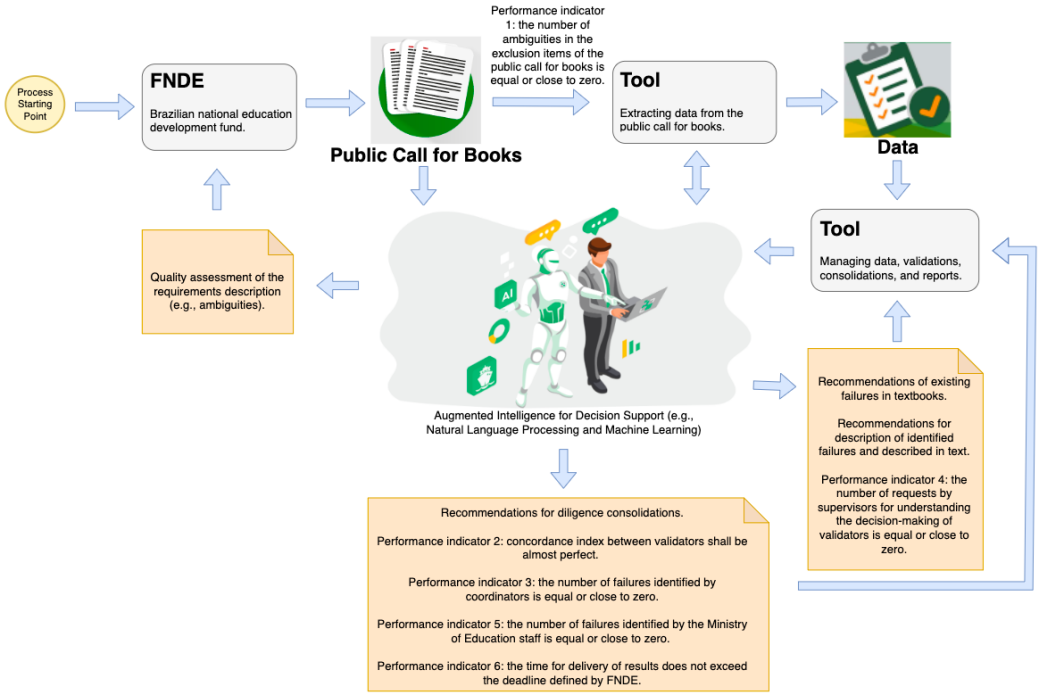

Figure 17. Overview of strategies and proposed performance indicators for automation category from our theory of change.

4.5.6. Presentation and Discussion with the FNDE

We organized a presentation and discussion with the FNDE aiming to validate and improve our theory of change. Based on the presentation, we evaluated the Ishikawa diagram and concluded that the proposed theory of change is in line with the FNDE's expectations for process improvements.

It is also essential to highlight that we are addressing some of the causes of problems by implementing the proposed strategies. Specifically, for the automation category, we

are addressing the following causes: lack of consistency in spreadsheets and difficulties in handling spreadsheets (e.g., generating reports). To address these issues, we are replacing spreadsheets with software for data management and reporting (automation strategy). Additionally, as mentioned in previous sections, we are also automating the validation of digital books using the software. In collaboration with the FNDE, we are already addressing some causes of problems related to the training and human resources category, such as incorrect understanding of the call for books and unplanned repeated diligence. Strategies include conducting training for validators to improve their understanding of the public call for books and providing tutorials. Furthermore, the FNDE is already addressing some of the causes of problems that this research cannot solve, such as low flexibility when using the platform and recurring platform problems (e.g., crashes). However, most of the issues highlighted in this article have not been addressed yet.

Therefore, this study presents some threats to validity. For instance, during our critical analysis, we did not consider the understanding of teachers' use of textbooks approved by the program. Additionally, we did not consider the opinions of public school students on the quality of approved textbooks.

5. Conclusions and Future Works

Ensuring quality education is a relevant goal for sustainable development. According to the SDGs report 2021 [34], for the global community, it is critical to ensure inclusive and equitable quality education and promote lifelong learning opportunities for all (goal 4). Such a goal comprises seven outcome targets: (4.1) universal primary and secondary education, (4.2) early childhood development and universal pre-primary education, (4.3) equal access to technical/vocational and higher education, (4.4) relevant skills for decent work, (4.5) gender equality and inclusion, (4.6) universal youth literacy, and (4.7) education for sustainable development and global citizenship. The PNLD supports the Brazilian government in achieving goal 4, highlighted in the SDGs report 2021, by improving the quality of books and other materials available for students and teachers and, consequently, improving the teaching–learning process in public schools and educational equity.

This study relates to target 4.1, which aims to ensure that by 2030, all students complete accessible, equitable, and quality primary and secondary education to achieve relevant and effective learning outcomes. Such a goal is crucial for low- and lower-middle-income countries, including Brazil, due to a large number of poor and vulnerable students in public schools. In addition to target 4.1, target 4.7 from goal 4 strongly relates to the impact of our study. In the Brazilian context, achieving goal 4 is even more challenging due to the country's continental size, socioeconomic inequalities, and regional differences. This study also relates to other goals, such as goal 9 (industry, innovation, and infrastructure—targets 9.5 and 9.b). The results presented in this study can indirectly support the Brazilian government in achieving such a goal. For instance, increasing the quality and accessibility of digital textbooks can assist the Brazilian government in increasing equity in Brazilian public schools and achieving these relevant SDGs.

Information and Communication Technologies (ICTs) can assist the government in achieving significant educational results, both at the administrative level and in the classroom. Decision support for public policy is crucial for improving the public sector [35]. However, the government must consider the Brazilian population's size, inequalities, and differences to appropriately implement ICT solutions [36].

Indeed, at the administrative level of public administration, decision making, defined processes, and appropriated ICTs can positively or negatively affect the quality of public education [37–39]. Specifically, adequate decision making plays a relevant role in the quality of public education. As an example of a concept for supporting decision making in public education, augmented intelligence-based strategies can integrate human agents and artificial intelligence in decision making, enabling intelligent decision support systems [33]. Furthermore, a good and modern internal organizational process is another factor that can ensure quality. The digital transformation of internal administrative processes is another

example of relevant action to enhance confidence in the public administration's success [7]. Plenty of evidence advocates for the relevance of implementing digitization in the public sector [8].

We identified and analyzed the causes of problems and their effects using an Ishikawa diagram developed based on interviews and internal documents. In addition, we proposed a theory of change based on the causes and effects, aiming to discuss the digital transformation of the validation and triage process for textbooks in the context of the PNLD. The Policy Design Arc framework of Harvard's Kennedy School of Government guided our methodological steps, providing a systematic and comprehensive mechanism for strategic thinking about the validation and triage process. We evaluated the Ishikawa diagram and the theory of change with stakeholders of the textbooks' validation and triage process and the FNDE.

Our study revealed problems and their underlying causes in the validation and triage process of textbooks subscribed to the PNLD program which negatively affect the quality of textbook delivery, including possible editing failures. Based on these findings, we propose the digital transformation of the validation and triage process using augmented intelligence techniques, such as machine learning, as a significant component of the theory of change strategies. Further studies could incorporate intelligent decision support systems into the process to address issues such as inconsistent analysis of textbooks, a high volume of manual tasks, conflicting validations, consolidation failures, and public calls for books inconsistence. In addition, in future works, it is also relevant to understand teachers' use of the textbook and students' opinions.

Author Contributions: Conceptualization, Á.S., I.I.B., A.C.M.d.S., A.P.d.S., D.D., L.B.M., N.C.I.R., A.C.S.e.S., R.F. and S.I.; Methodology, Á.S., I.I.B., A.C.M.d.S., A.P.d.S., D.D., L.B.M., N.C.I.R., A.C.S.e.S., R.F. and S.I.; Investigation, Á.S., I.I.B., A.C.M.d.S., A.P.d.S., D.D., L.B.M., N.C.I.R., A.C.S.e.S., R.F. and S.I.; Writing—original draft, Á.S., I.I.B., A.C.M.d.S., A.P.d.S., D.D., L.B.M., N.C.I.R., A.C.S.e.S., R.F. and S.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by FNDE and the Brazilian Ministry of Education, Brazil, Decentralized Execution Term (Termo de Execução Descentralizado—TED) 10320—Inteligência Aumentada para Validação e Análise de Atributos do PNLD.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zambon, L.B.; Terrazzan, E.A. Policies of didactic material in Brazil: Organization of procedures for selection of textbooks in basic education public schools. *Rev. Bras. Estud. Pedagog* **2013**, *94*, 585–602. [CrossRef]
2. Shinde, D.D.; Ahirrao, S.; Prasad, R. Fishbone Diagram: Application to Identify the Root Causes of Student–Staff Problems in Technical Education. *Wirel. Pers. Commun.* **2018**, *100*, 653–664. [CrossRef]
3. Reinholz, D.L.; Andrews, T.C. Change theory and theory of change: What's the difference anyway? *Int. J. STEM Educ.* **2020**, *7*, 2. [CrossRef]
4. Filgueiras, F.; Flávio, C.; Palotti, P. Digital Transformation and Public Service Delivery in Brazil. *Lat. Am. Policy* **2019**, *10*, 195–219. [CrossRef]
5. Ahn, M.J.; Chen, Y.C. Digital transformation toward AI-augmented public administration: The perception of government employees and the willingness to use AI in government. *Gov. Inf. Q.* **2022**, *39*, 101664. [CrossRef]
6. Datta, P.; Walker, L.; Amarilli, F. Digital transformation: Learning from Italy's public administration. *J. Inf. Technol. Teach. Cases* **2020**, *10*, 54–71. [CrossRef]
7. Scupola, A.; Mergel, I. Co-production in digital transformation of public administration and public value creation: The case of Denmark. *Gov. Inf. Q.* **2022**, *39*, 101650. [CrossRef]
8. Giulio, M.D.; Vecchi, G. Implementing digitalization in the public sector. Technologies, agency, and governance. *Public Policy Adm.* **2021**, 09520767211023283. [CrossRef]

9. de Mattos Höfling, E. Notes for discussion about the implementation of government programs: focusing on the Textbook National Program. *Educ. Soc.* **2000**, *21*, 159–170. [CrossRef]
10. Bianco, A.A.G. Análise do conteúdo imagético de nutrição humana em livros didáticos de Ciências aprovados pelo Programa Nacional do Livro Didático 2014. *ABCS Health Sci.* **2015**, *40*. [CrossRef]
11. de Souza, L.H.P.; Rego, S.C.R. Images in science textbooks and guidelines from Brazil's national textbook program. *Ensaios Pedagógicos* **2018**, *2*, 5–15.
12. Manoel, C.C.; Silva, M.; Valero, P. Happy and healthy families! Financial mathematics and the making of the homus oeconomicus. In Proceedings of the Tenth International Mathematics Education and Society Conference, Hyderabad, India, 28 January–2 February 2019.
13. de Albuquerque, E.B.C.; Ferreira, A.T.B. Programa nacional de livro didático (PNLD): Mudanças nos livros de alfabetização e os usos que os professores fazem desse recurso em sala de aula. *Ens. Avaliação E Políticas Públicas Em Educ.* **2019**, *27*, 250–270. [CrossRef]
14. Wahlin, W.E. The Social is the Thing: Taking a Designerly Approach to Local Government Strategic Planning and Evaluation. In Proceedings of the Sixteenth International Conference on Design Principles and Practices: Back to Life, Newcastle, Australia, 19–21 January 2022.
15. Carvalho, R.; Lobo, M.; Oliveira, M.; Oliveira, A.R.; Lopes, F.; Souza, J.; Ramalho, A.; Viana, J.; Alonso, V.; Caballero, I.; et al. Analysis of root causes of problems affecting the quality of hospital administrative data: A systematic review and Ishikawa diagram. *Int. J. Med. Inform.* **2021**, *156*, 104584. [CrossRef]
16. Idris, N.I.; Sin, T.C.; Ibrahim, S.; FadzliRamli, M.; Ahmad, R. A Case Study of Coffee Sachets Production Defect Analysis Using Pareto Analysis, P-Control Chart and Ishikawa Diagram. In *Lecture Notes in Mechanical Engineering*; Springer: Singapore, 2021; pp. 1295–1305. [CrossRef]
17. Sharma, G.V.S.S.; Prasad, C.L.V.R.S.V.; Rambabu, V. Online machine drawing pedagogy—A knowledge management perspective through maker education in the the COVID-19 pandemic era. *Knowl. Process Manag.* **2021**, *29*, 231–241. [CrossRef]
18. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46. [CrossRef]
19. Landis, J.R.; Koch, G.G. A Coefficient of Agreement for Nominal Scales. *Biometrics* **1977**, *33*, 159–174. [CrossRef]
20. McHugh, M.L. Interrater reliability: The kappa statistic. *Biochem. Med.* **2012**, *22*, 276–282. [CrossRef]
21. Anderson, A. *The Community Builder's Approach to Theory of Change: A Practical Guide to Theory Development*; Aspen Institute Roundtable on Community Change: New York, NY, USA, 2005.
22. Weiss, C.H. Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. *New Approaches Eval. Community Initiat. Concepts Methods Context* **1995**, *1*, 65–92.
23. Gilissen, J.; Pivodic, L.; Gastmans, C.; Stichele, R.V.; Deliens, L.; Breuer, E.; den Block, L.V. How to achieve the desired outcomes of advance care planning in nursing homes: A theory of change. *BMC Geriatr.* **2018**, *18*, 47. [CrossRef]
24. Pittock, J.; Bjornlund, H.; van Rooyen, A. Transforming failing smallholder irrigation schemes in Africa: A theory of change. *Int. J. Water Resour. Dev.* **2020**, *36*, S1–S19. [CrossRef]
25. Lam, S.; Dodd, W.; Wyngaarden, S.; Skinner, K.; Papadopoulos, A.; Harper, S.L. How and why are Theory of Change and Realist Evaluation used in food security contexts? A scoping review. *Eval. Program Plan.* **2021**, *89*, 102008. [CrossRef]
26. Davenport, C.; Dele-Ajayi, O.; Emembolu, I.; Morton, R.; Padwick, A.; Portas, A.; Sanderson, J.; Shimwell, J.; Stonehouse, J.; Strachan, R.; et al. A Theory of Change for Improving Children's Perceptions, Aspirations and Uptake of STEM Careers. *Res. Sci. Educ.* **2020**, *51*, 997–1011. [CrossRef]
27. Briz-Redón, Á. Respondent Burden Effects on Item Non-Response and Careless Response Rates: An Analysis of Two Types of Surveys. *Mathematics* **2021**, *9*, 2035. [CrossRef]
28. Ain, N.; Vaia, G.; DeLone, W.H.; Waheed, M. Two decades of research on business intelligence system adoption, utilization and success—A systematic literature review. *Decis. Support Syst.* **2019**, *125*, 113113. [CrossRef]
29. Sternberg, J. *Beyond Spreadsheets*; Forbes Media LLC: New York, NY, USA, 2020.
30. Saxena, D.; McDonagh, J. Communication breakdowns during business process change projects—Insights from a sociotechnical case study. *Int. J. Proj. Manag.* **2022**, *40*, 181–191. [CrossRef]
31. Bana, S.H.; Benzell, S.G.; Solares, R.R. Ranking How National Economies Adapt to Remote Work. *MIT Sloan Management Review*, 18 June 2020.
32. Rabiee, M.; Aslani, B.; Rezaei, J. A decision support system for detecting and handling biased decision-makers in multi criteria group decision-making problems. *Expert Syst. Appl.* **2021**, *171*, 114597. [CrossRef]
33. Toivonen, T.; Jormanainen, I.; Tukiainen, M. Augmented intelligence in educational data mining. *Smart Learn. Environ.* **2019**, *6*, 10. [CrossRef]
34. United Nations. *The Sustainable Development Goals Report 2021*; United Nations: New York, NY, USA, 2021.
35. Marchi, G.D.; Lucertini, G.; Tsoukiàs, A. From evidence-based policy making to policy analytics. *Ann. Oper. Res.* **2014**, *236*, 15–38. [CrossRef]
36. Valente, J.A.; de Almeida, M.E.B. Políticas de tecnologia na educação no Brasil: Visão histórica e lições aprendidas. *Educ. Policy Anal. Arch.* **2020**, *28*, 94. [CrossRef]
37. Ferman, B.; Finamor, L.; Lima, L. *Are Public Schools in Developing Countries Ready to Integrate EdTech into Regular Instruction*? Munich Personal RePEc Archive: Munich, Germany, 2019.

38. Styles, B.; Torgerson, C. Randomised controlled trials (RCTs) in education research—Methodological debates, questions, challenges. *Educ. Res.* **2018**, *60*, 255–264. [CrossRef]
39. Escueta, M.; Nickow, A.J.; Oreopoulos, P.; Quan, V. Upgrading Education with Technology: Insights from Experimental Research. *J. Econ. Lit.* **2020**, *58*, 897–996. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.