Lesson plan analysis protocol (LPAP): A useful tool for researchers and educational evaluators

Kizito Ndihokubwayo*, Céline Byukusenge, Edwin Byusa, Hashituky Telesphore Habiyaremye, Agnes Mbonyiryivuze, Josiane Mukagihana

African Centre of Excellence for Innovative Teaching and Learning Mathematics and Science (ACEITLMS), University of Rwanda College of Education (URCE), Kayonza, PO BOX 55, Rwamagana, Rwanda

ARTICLE INFO

Keywords:
Analysis protocol
Competence-based education
Lesson plan
Pedagogical document

ABSTRACT

In improving learning outcomes in education, we found a gap in available standard protocols to analyze pedagogical documents such as lesson plans. This study is a product of validated and reliable Lesson Plan Analysis Protocol (LPAP) supporting education stakeholders to get insight into the lesson plans (LPs) used in schools. The LPAP was found to be a valid and reliable tool that education evaluators can use to strengthen effective teaching in all grades of education. The protocol can also serve as teacher self-evaluation before lesson delivery. After validating it, we collected lesson plans to test its usability and a qualitative survey for its inclusivity. We analyzed 36 among many collected lesson plans in two stages. We showed how to analyze averages for each of 27 LPAP items and how to analyze the mean scores for each lesson plan. We again collected reflection information from six researchers and found them appreciating the tool. Thus, LPAP is a useful tool to analyze taught or untought lesson plans from any country and with any lesson plan format as it reserved some flexibilities to be modified to fit the context.

1. Introduction

The Ministry of Education (MINEDUC) in Rwanda, through the Rwanda Basic Education Board (REB), has embarked on reviewing its curriculum named the Competence-Based Curriculum (CBC) in 2015 to align it with national aspirations (REB, 2015). This curriculum reform was done to ensure that the knowledge, skills, attitudes, and values acquired by Rwandans in schools meet the challenges of the 21st century. The introduction of this curriculum requires large-scale efforts to prepare the education system to deliver its approaches rapidly. "Curriculum change on its own to bring about transformation is incomplete without a simultaneous change in pedagogy (Blignaut, 2020, p. 1)". Therefore, action is needed to push the status from the current level to the expected level in all aspects of CBC understanding, implementation, and outcomes by starting with the lessons' preparation.

According to CBC concepts, the learning process is learner-focused, where a learner is engaged in active and participatory learning activities. The learner builds new knowledge from prior knowledge through discovery and problem-solving-based learning (Mbarushimana and Kuboja, 2016). All these processes of learning must be explicit in lesson plan steps. For teachers to support learners, the assessment happens as an integral part of the learning process. It takes place through informal or formal methods set in each step of the lesson plan, from introduction to the lesson's closure. This assessment is mainly criterion-referenced to evaluate and measure what learners can demonstrate in their learning. Consequently, if the teacher cannot prepare a good lesson plan, learning will not teach the CBC lesson. For instance, in the third phase of CBC training (JICA, 2020; REB, 2018), teachers were showed three consecutive phases of the lesson, and each one has its role in delivering an effective lesson. These phases are pre-planning, planning, and lesson delivery phases.

In the pre-planning phase, the teacher thinks about how, what, when, and to whom the lesson will be taught. This first phase is a planning process—when the teacher feels about what will be taught. It is essential as it plans the creativity finally seen in the delivered lesson. The second phase is the planning phase, in which what was planned in the teacher's mind is written on paper or is filled in the lesson plan form. Furthermore, this is the case with a typical lesson plan. All necessary pedagogical documents (such as syllabus, the scheme of work, teacher's guide, and student's textbook), teaching strategies (such as roleplay, individual
learning, group work, questions in the corner), teaching aids (such as lab materials, ICT equipment, improvised materials), and inputs (such as field trips, expert visits) are put together. The last phase is the lesson delivery, where the planned lesson is implemented in a classroom. During lesson delivery, learners are taught and guided. They are engaged, assessed, and therefore gain competences.

In partnership with various developmental partners (DPs) such as Japan International Cooperation Agency (JICA)—through its piloting project for supporting institutionalizing the quality of SBI1 activity (SIQoS, JICA, 2020), Building Learning Foundation (BLF), British Council, Wellspring, Educate! Inspire-Educate-Empower (IEE-Rwanda), to mention few, REB has continuously assisted and trained teachers in implementing CBC across the country. They also frequently visit schools and observe classrooms to track teaching shifts from teacher to learner-centeredness. The classroom observation and evaluation cannot be done separately from the classroom activities to the planned activities. The observers have to assess the lesson plans before observing the class. We have evaluated some observation sheets (such as REB, 2019 pp. 89–93); however, there was no standardized tool to assess the lesson plan (LP) before classroom observation. This gap necessitated a tool that could be used to assess the lesson plans for competence-based compliance before they are taught. This protocol is of interest to teachers in various subjects as they would prepare a clear and standard LP that everyone can use in absence of the owner—the one who prepared it. This tool is very crucial and essential for teachers. It eases school administrators’ work due to the same content of CBC leading to the same lesson plan. More information on this protocol is provided in the methodological part and Training manual in supplementary materials.

2. Analysis of LP format across African countries

We have analyzed the LP formats across African countries before developing the new protocol. We have also so far consulted about 5 lesson formats across African countries (see Table 1).

There are differences in the lesson plan format used in African countries. The main difference is the realization of the curriculum used. Wherever information of competences in the lesson plan is found, one may guess that the country is implementing the competence-based curriculum.

3. Reviewed existing lesson plan analysis tools

Pedagogical documents are essential for teachers. These documents guide teachers (Njiku, 2016) and form a coherent unit describing individualized and structured education plans. These plans should focus on a particular learner and take the child’s individuality into account in education and support planning (Heiskanen, 2019). A lesson plan is a blueprint of teaching practice. It is an activity conducted by a teacher before implementing it in the classroom during teaching and learning processes, fitting learners’ needs (Raval, 2013). A lesson plan plays a role in the education system. For instance, it serves as the key to students’ achievement and teachers’ attitudes (Nesari and Heidari, 2014). Thus, the planning reflects the teacher’s real practices on the ground or application into the entire classroom. The lesson consists of a unit of analysis, reducing teaching complexity to a manageable size without altering its character (Santagata et al., 2007). The essential elements of any instructional effort, such as goals for students’ learning, instructional activities, strategies for monitoring students’ thinking, and assessing their learning, curriculum, pedagogy, and so forth, are included in the lesson (Santagata et al., 2007). While planning and teaching through daily lessons, most teachers reflect on all these elements. Lesson planning involves teachers’ decisions related to lesson preparation (Taylan, 2016). Different studies on designing tools for analyzing teachers’ lesson plans from other countries worldwide have been conducted, and some of them are highlighted in this section.

In addressing an issue of a need for pedagogical tools that help teachers develop essential pedagogical content knowledge and practices to meet the mathematical education needs of a growing culturally and linguistically diverse student population in the USA, researchers introduced an innovative lesson analysis tool focusing on culturally responsive mathematics teaching (CRMT) (Aguirre and Zavala, 2013). The tool focuses on integrating mathematical thinking, language, culture, and social justice. This study showed that the tool enabled teachers to systematically analyze and critique mathematics lessons with multiple dimensions, including mathematical thinking, language, culture, and social justice.

Ferrell (1992) from the University of Texas Medical Branch designed a Lesson Plan Evaluation Form (LPEF) that can be used to provide systematic quantitative data about classroom functioning that are usually obtained only through attitudinal surveys. The author argued that the LPEF could be used to provide information to program decision-makers that could not be obtained through the usual evaluation procedures and provided for more detailed documentation of data usually obtained through attitudinal assessments. Moreover, the tool was claimed unable to determine whether the program is implemented according to program guidelines or not.

Jacobs et al. (2008) developed and validated the Science Lesson Plan Analysis Instrument (SLPAI). It is an excellent analysis tool aiming to evaluate teacher development programs’ success for quantitative evaluation of teacher-generated multiday lesson plans. It was also used to track changes in teaching practice and pedagogical knowledge of participants over time and for providing summative evidence of program effectiveness.

The purpose of the Goldston et al. (2010) study was to describe the procedures and the analysis of an instrument designed to measure pre-service teachers’ ability to develop appropriate 5E learning cycle lesson plans. The authors developed and validated the 5E inquiry lesson plan (ILP) rubric that comprises 12 items with a scoring range of zero to four points per item.

Despite the available designed and validated Lesson Plan Analysis tools, none suits evaluation of competence-based education. Therefore, there was a need to develop a new analysis tool suitable to evaluate a—daily lesson plan practice such as one used in Rwanda or other countries that are currently implementing the competence-based curriculum. Thus, this study was conducted to develop a validated and reliable tool supporting education stakeholders to get insight into the lesson plan (LP) used in schools.

4. Methodology

4.1. Initial development of the protocol

The development of the LPAP was motivated by the need to have a protocol to analyze the competence-based pedagogical lesson plan. It started with the name “Analysis of Lesson Plan Protocol (ALPP)” as a checklist for teachers, policymakers, education evaluators, education developmental partners, and researchers. The process of ALPP development involved comprehensively consultation of the literature and educational policy documents, accurately REB documents. The ALPP had 13 items within six groups with different grades or ranking styles. These six groups and their 13 corresponding items are presented in Table 2.

The ranking categories varied from three categories (e.g., Not written, Unclear, and Well described for the instructional objective, IO or description of teaching and learning activity, DTLA) to seven categories (e.g., Single title, Double Title, Triple (or more than two) title, Time-bound, Scheme of Work connected, textbook connected, and teacher’s guide connected for the Title and its connection). The ALPP total score was 42 with the interpretation of poor lesson plan score below 21 out of 42 (below 50%), good lesson plan scored between 21 and 31.5 out of 42.

---

1 SBI: School-Based In-service Teacher Training.
4.2. Validation of LPAP

We sent the protocol to different stakeholders and counterparts such as REB staff, University lecturers who took part in CBC development, DPs, postgraduate students, and teachers for content validity check-ups. (a) One University Lecturer. He is a URCE retired chemistry lecturer. He has been a part of CBC development, and he has experience in developing and training teachers at a university level. (b) One REB staff. He works in the Teacher Development and Management and Career Guidance and Counselling department (TDM & CGC). He got enough experience from teachers’ training activities. (c) Two developmental partners (DPs). One DP staff is from IEE-Rwanda, while another is from the WellSpring foundation of education Rwanda. Both NGOs participated in various teacher training on Rwandan CBC implementation. (d) Five in-service secondary teachers. These are TTC (Teacher training college) teachers and NTs (national trainers). They hold an excellent position to validate our tool as they got experience in training primary teachers and train other fellow teachers on implementing CBC at the national level. (e) Six postgraduate students. Five students are Ph.D. while one is a Master’s student at the University of Rwanda College of Education (URCE), specifically from the African Center of Excellence for Innovative Teaching and Learning Mathematics and Science (ACEITLMS). Most of them have joined the center from a teaching career. Since they were teachers and are currently doing research in education, they fit our tool’s great validators.

We shared with them the lesson plan analysis protocol (LPAP) with its training manual and procedure for validation. We asked all validators to do the following: (a) To download the REB lesson plan format, (b) To use the LPAP document and write their comments or suggestions down, (c) To check if all crucial elements in the lesson plan were exhausted in the LPAP, (d) To suggest additions, modifications in rating scales (such as not written, unclear, well described,...), (e) To criticize the proposed scoring scheme, (f) To criticize the proposed interpretation of the results (poor as under 50%, good lesson plan as between 50 to 75%, and excellent as up to 100%), and (g) To provide any other suggestions or any questions related to the protocol or its scoring scheme.

4.3. Final version of LPAP

Based on validators’ reports, their suggestions and inputs were used to revise our protocol and get the updated and current version (see Appendix A in Training manual of supplementary materials). For instance, we removed some related components, such as the link between the lesson title and Scheme of Work/Textbook/Teacher’s guide. We added two new components (Key unit competence and lesson approaches) due to their role in showing what will be done to develop the expected skills and competence and develop curiosity among the learners. Lesson description (DTLA) was separated from instructional objectives to make another component. After compromising on the point, that lesson content does not reflect the stages of the lesson (Intro, Dev, and Concl), component “Lesson content” has been replaced by “lesson stages.” However, some components present in the REB lesson plan format are not included in this protocol—such as “plan for this classroom” and “references”—not because they are less important, but they were out of our research interest.

Table 1. Analysis of lesson plan formats across Africa.

| Preliminary information | Lesson body | Self-evaluation |
|------------------------|-------------|-----------------|
| Rwanda | Teacher fills information related to key unit competences of the lesson, instructional objectives, and special education needs | The teacher outlines and describes his/her and learners’ activities. He/she mentions and describes which generic competences learners should get and possible cross-cutting issues that should be catered for during the lesson | The teacher needs to describe how the lesson went and a possible way forward |
| Burundi | Teacher state objectives | Teacher outlines his/her and learners activities | There is no self-evaluation. However, on the right corner (after learners’ activities), there is an observation to be filled by the inspector |
| Uganda | The teacher fills the information related to generic competences, learning outcomes, values, and cross-cutting issues | All three parts are tackled in phases presenting teacher and learners activities. For learners, there is discovery, explanatory, analysis, and application | The teacher needs to describe how the lesson went |
| Tanzania | The teacher needs to state competence and lesson objectives | There are three phases. Development of knowledge, application, and reflection | In addition to teacher self-evaluation, there are also learners evaluation and remarks |
| Malawi | The teacher needs to state the success criteria. This is like lesson objectives | The teacher outlines what needs to be learned. This part may have steps but are not standards named | Instead of teacher self-evaluation, there is lesson evaluation. |

Table 2. Initial development of LPAP.

| Groups | Items |
|--------|-------|
| 1      | Title and its origin | Title and its connection |
| 2      | Instructional objective and lesson description | Instructional objective |
|        | Description of teaching and learning activity |
| 3      | Special education needs | Special education needs (SEN) |
| 4      | Lesson content | Introduction to the lesson (Intro) |
|        | Lesson development (Dev) |
|        | Conclusion of the lesson (Concl) |
|        | Teaching resources (TR) |
|        | Active learning techniques (ALT) |
|        | Formative assessment (FA) |
| 5      | Cross-cutting issues and development of competences | Cross-cutting issues (CCs) |
|        | Generic competencies (GCs) |
| 6      | Lesson evaluation | Teacher self-evaluation (TSE) |

(50–75%), and excellent lesson plan scored above 31.5 out of 42 (above 75%). The protocol got different updates through the consultation with all authors, including the name’s changes to “Lesson Plan Analysis Protocol (LPAP).”
The produced final protocol has nine groups of 27 items with different grading or ranking. This is provided in the training manual (see Appendix B in the Supplementary materials). These nine groups are within three stages: preliminaries, the body of the content, and the accessories. Despite their differences in wording, all the items in the final version have four scales and measure the same weight. For instance, the first two scales or answer categories should be scored zero; the third scale should be scored one score, while the last scale or the fourth category should be marked two scores. Regarding LPAP data analysis, the preliminary groups have 18 scores, the body of the content has 30, while accessory groups have six scores. Therefore, the total scores of the final version of LPAP are 54, and their respective interpretations range from a poor, fair, good, very good, and excellent lesson plan. Note that the validators helped us to form these ranging figures. Depending on the scores on a specific lesson plan, it can be recommended to be taught. Thus, the lesson plan having a score (a) below 27 out of 54 (below 50%) is considered as a poor lesson plan, and it cannot be taught, (b) between 27 and below 37.8 out of 54 (50–69%) is considered as a fair lesson plan, and it cannot be taught, (c) between 37.8 and below 43.2 out of 54 (70–79%) is considered as a good lesson plan and it can be taught, (d) between 43.2 and below 48.6 out of 54 (80–89%) is considered as a very good and it can be taught, and (e) between 48.6 and above out of 54 (90–100%) is considered as an excellent and it can be taught.

4.4. Reliability testing

Upon getting good content validity, we sampled two lesson plans and rated them to seek interrater reliability. We used MS Excel 2016 to compute the agreement between every two raters/coders. Each rater could rate each of 27 LPAP items into one of its four scales (1, 2, 3, or 4). Then, the difference between scores from two raters is computed along with each LPAP item. After, we count a number of zeros—where no difference occurred (or where both raters coded similarly). The percentage of this count produces the agreement between these two raters. The authors rated the first lesson plan, and the highest agreement of one of the pairs was 54%. At this stage, the authors were still practicing the protocol.

The modification was done due to some challenges found in some items, such as lesson stages and approaches. Moreover, some raters were commenting inside the boxes instead of the provided space under "comment" boxes. Thus, the protocol has been revised to cater to those issues. To attain a high coefficient, we improved the protocol by bringing together all the items with an equal scaling system to calculate reliability and become friendly. Despite their differences in wording, all the final version items have four scales and measure the same weight. For instance, the first two scales or answer categories should be scored zero; the third scale should be scored one score, while the last scale or the fourth category should be marked two scores. The training manual (see Supplementary materials) was also refined based on the protocol's improved version.

Then, all authors have redone the rating with another lesson plan for a different subject. This time, we computed Cohen's kappa reliability to minimize the ratings by chance, using SPSS version 23.0 under the "Descriptive Statistics" function [Analyze > Descriptive Statistics > Crosstabs] and select Kappa statistics under 'Statistics' option.

The second rating for reliability generated 74% agreement with the .877 coefficient of Spearman correlation and .640 coefficient of Kappa statistics. The highest agreement (85%) was found in pair of first and fourth coders, while the lowest agreement (67%) among six authors/coders was found between coders 5 and 6 (see Figure 1). Such proportional agreement among more than two raters would be appreciated (Fleiss, 1971; Krippendorff, 2011). The intense level of agreement ranges from .80 to .90, and this implies that 64–81% of data are reliable (Berry and Mielke, 1988; De Vries et al., 2008; McHugh, 2012).

5. Testing LPAP

After validating our tool in 2020, we then went ahead to test it. We have collected some lesson plans in 2021 for this purpose. We asked teachers via WhatsApp groups to share with us any used lesson plans. We got responses within one week. We received various lesson plans from various subjects and various teachers across the country. Some lesson plans were more than ten from the same teachers, others were more than twenty from the same subjects we got lesson plans from different teachers, and some teachers provided lesson plans from different subjects. However, we got single lesson plans; thus, some subjects got only one or two lesson plans from one or more teachers. Therefore, we listed all lesson plans we received according to teachers and subjects and successfully classed nine subjects. These subjects got at least four lesson plans from one teacher. Where we saw more than one teacher in one subject, we randomly selected one teacher. Thus, we took 36 lesson plans from nine subjects—or from nine teachers since each four lesson plans of one subject were from one teacher—to the analysis phase.

Figure 1. Intercoder reliability.
Among the lesson we selected, English was planned from 22 May to July 2019, Physics from 03 to 06 March 2020, Biology from 12 to 13 September 2020, Geography from 23 November to 02 December 2020, Entrepreneurship from 28 January 2020 to 24 May 2021, History from 24 March to 04 May 2021, Mathematics from 21 April to 19 May 2021, and Kiswahili and Kinyarwanda did not specify data.

We used an Excel spreadsheet to enter data and analyzed data. Three authors and three external raters rated or coded data using the form in appendix A. These external raters were master students at URCE and have participated in the validation of LPAP. Thus, they are familiar with LPAP. However, they went into reliability check-ups, and their agreement was above 80%, while Kappa statistics were above 0.7. Each author and external rater coded 12 LPs (see data in the Supplementary materials). The data from the LPAP form were transferred to Excel software. Data were entered so that each scale (among four) coded, its corresponding number was entered. For instance, if the author or rater coded the first LPAP in the fourth scale, “4” was entered. Likewise, if the author or rater finds the teacher did not write anything in the 27th LPAP item (teacher self-evaluation), “1” was entered in the software. We analyzed the data in two stages.

5.1. Analysis 1: analysing LPAP among its 27 elements

We, in the first stage, averaged the codings from the author and external rater. Then we computed average and standard deviations along with all lesson plans for each LPAP item. For more detail, see data in the Supplementary materials. Table 3 presents the average and standard deviations across all 27 LPAP items. Note that the highest assigned number an item would have is “4” and 1 would be “1.” Therefore, the number close to four indicates the occurrence and rightfulness of the item.

Teachers were found writing well the key unit competence (KUC) and have a good practice of copying it from the syllabus. This is shown by 3.72 out of 4 scales. The title of the lesson was in a good format, time-bound, and connected to the syllabus. Instructional objectives (IO) were well set, although its components (conditions, who, action, content, and standard of performance) were not all covered. Some teachers miss out on the condition and standard of performance. However, the condition, such as materials to be used is important as it shows what is needed to reach the objectives. Likewise, the standard of performance is crucial as it serves as a criterion to measure and evaluate whether the objective was achieved or not. Most teachers leave empty or just write “none” in the space reserved for special education needs (SEN), and eventually, it was difficult to cater for it in the body of the lesson (1.18 out of 4) as they did not plan for it. The content in the lesson stages is outlined but not fully described, and none use the format in appendix C (see Training manual in supplementary materials). The reason may be that some schools provide lesson plan formats with limited space, and therefore, a description of activities may not be possible.

Lesson approaches such as active learning techniques (ALT), teaching resources (TR), and formative assessment (FA) were problematic (most were coded in the first and second scales). Thus, teachers do not plan for or show which techniques, teaching aids they will use, or how they will assess the learning. They present cross-cutting issues (CCIs) and generic competences (GCs), although they do not describe how they would be addressed and attained, respectively. Teachers were not eager to fill in the space reserved for their self-evaluation (2.35 out of 4). They left it empty or just wrote: “done.” This would be a good opportunity to evaluate his teaching and show a way forward.

5.2. Analysis 2: analyzing whether lesson plan can be taught or not

In the second stage, we transformed the averaged codings from the author and external rater into scores as suggested in Appendix B (see Training manual in supplemental materials). Since LPAP has 27 elements or items and is coded into four scales, the first and second scales are given a “0” score, the third scale is given a “1” score, while the fourth scale is given “2” scores. Thus, the maximum score an LPAP item would get is “2.” Therefore, the total score is 54 (see methodology section for more detail). This is to say that where we had “1” or “2” in the above analysis, we got zero, and where we had “3”, we got one, while where we had “4”, we got two. We then computed the sum for each lesson plan along with each LPAP item. We computed the percentage from this sum, and Figure 2 was plotted.

We found that all lesson plans across all subjects/teachers scored below 50%. Mathematics scored high among other subjects (between 40% and 50%) while Entrepreneurship scored lowest (between 10% and 20%). Note that these data demonstrate how to use and analyze LPAP data; otherwise, we would conclude that none among 36 collected, coded, and analyzed lesson plans would be taught. All lesson plans are considered as poor as they got below 50% score (below 27 out of 54), and they cannot be taught. Being taught, we refer to whether such a lesson plan can help someone who did not prepare it to deliver it. The fact that we did not find any lesson plan that qualifies for being taught, such as a good lesson plan (70–79%), maybe depicted from the purpose of these data and sampling technique used.

6. Reflection on LPAP

We have also, after development, approached researchers for qualitative feedback. We used the usual ACETILMS WhatsApp group and asked for anyone, volunteerily, to provide us any comment, appreciation, or critical to LPAP. We shared the manuscript and Training manual. We
received feedback from six students—three Ph.D. students from Tanzania, Liberia, and Zambia; and three master students, one among them a female from Rwanda. All appreciated the tool in its development and usability.

Thank you very much for this wonderful instrument. In my view, it is an innovative and commendable move that you have made. I hope it is completed soon so that we can adapt it to our various contexts.

The protocol is good, I can use it as a researcher, and it is a good tool to be used by anyone, even in an international context, since the teaching practices are almost the same across the countries. However, some improvements are on the scoring; it is not fair that scale 1 and scale 2 have all 0 scores.

I appreciate this tool as well as the author who prepared it. It is well prepared, and it is well understood for everyone after reading the training manual and can be adopted by the researcher to assess any lesson plan not only for science subject but for all subject taught in any school. However, you mentioned that an observer (Rater) should tick with \( \checkmark \) or mark with “Yes, but while completing (Appendix: B. Lesson Plan Analysis Protocol (LPAP) Scoring Scheme) table across each item you rated them using 0, 1, and 2.

First of all, I appreciate your hard-working spirit in developing this tool; it was really needed in education. It is well planned and clear to be used. I am very interested in its ability to be applied in all subjects. However, the new format is not provided, and no teacher can imagine putting these components.

7. Conclusion and mode of LPAP usability

The study’s design was motivated by the gap identified in the lack of tools to analyze pedagogical documents such as lesson plan that reflect on the competency-based pedagogy. The draft protocol was developed first, together with its training manual (see Training manual). The produced initial protocol was sent to different experts with considerable experience in the competence-based curriculum for validity purposes; this was done and yielded an improved version of the LPAP. After this stage, the reliability check process started; through this process, a very good LPAP was produced. The LPAP is a valid and reliable tool for teachers and educational evaluators. It can still be extrapolated to any other countries’ curriculum because the researchers or evaluators may modify some of the LPAP items to fit their purpose and curriculum in use.

LPAP may be used in the following modes:

(a) since we found that teachers are not yet adapt to the new lesson plan format (as in appendix C), item-13 and 14 may be omitted. This may be applicable to evaluators outside Rwanda, who may find some irrelevant or inappropriate items. The items may be omitted or modified, and remember to adjust scores in Appendix B.

(b) One may choose to code or rate lesson plans using the LPAP form in appendix A or an Excel sheet (data entry form in Supplementary material).

(c) One may choose to analyze data using one or both approaches. Analysis one that analyses scales (maximum will be “4” while the minimum will be 1) provides a room to analyze each item on LPAP, which practices teachers perform well or badly while preparing the lesson. Analysis two that analyses scores (maximum will be “2” while the minimum will be 0) gives room to analyze the subject, teachers, or a single lesson plan, then you can conclude whether or not a certain lesson plan deserves to be taught or whether one can leave it to a colleague to use.

(d) LPAP may be taken into two considerations. The first way is to consider a taught lesson plan, and in this case, teacher self-evaluation (TSE) or 27th LPAP item would be omitted (not be considered).

Declarations

Author contribution statement

Kizito Ndihokubwayo: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.
Céline Byukusenge and Josiane Mukagihana: Conceived and designed the experiments; Performed the experiments; Wrote the paper.
Edwin Byusa, Telesphore Hasituky Habiyaremye and Agnes Mboniyivu: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.
**Funding statement**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Data availability statement**

Data included in article/supplementary material/referenced in article.

**Declaration of interests statement**

The authors declare no conflict of interest.

**Additional information**

Supplementary content related to this article has been published online at [https://doi.org/10.1016/j.heliyon.2022.e08730](https://doi.org/10.1016/j.heliyon.2022.e08730).

**Acknowledgements**

We are grateful to everyone who participated in this study. We would like to pay our heartfelt feeling to people that validated this tool, supplied their countries lesson plan formats, supplied qualitative reflection, and commented on the manuscript or its training manual. We highly acknowledge the teachers who provided their lesson plans.

**References**

Aguirre, J.M., Zavala, R., 2013. Making culturally responsive mathematics teaching explicit: a lesson analysis tool. Pedagogies: Int. J. 8 (2), 163–190.

Berry, K.J., Mielke, P.W., 1988. A generalization of Cohen’s kappa agreement measure to interval measurement and multiple raters. Educ. Psychol. Meas. 48 (4), 921–933.

Goldston, M.J., Day, J.B., Sundberg, C., Dantzler, J., 2010. Psychometric analysis of a 5e learning cycle lesson plan assessment instrument. Int. J. Sci. Math. Educ. 8, 633–648.

Jacobs, C.L., Martin, S.N., Otieno, T.C., 2008. Instrument for formative and summative program evaluation of a teacher education program. Sci. Teach Educ. 92, 1097–1126.

Krippendorff, K., 2011. Communication methods and measures agreement and information in the reliability of coding. Commun. Methods Meas. 5 (2), 93–112.

McHugh, M.L., 2012. Lessons in biostatistics interrater reliability: the kappa statistic. Biochem. Med. 22 (3), 276–282. [https://hrcak.srce.hr/89395](https://hrcak.srce.hr/89395).

Raval, D.K., 2013. Lesson plan: the blueprint of teaching. Int. J. Relig. Educ. 2 (2), 155–157.

Taylan, R.D., 2016. The relationship between pre-service mathematics teachers’ focus on student thinking in lesson analysis and lesson planning tasks. Int. J. Sci. Math. Educ. 16 (2), 337–356.