Selecting and designing formal assessments of student learning

What is the practice of selecting and designing formal assessments of student learning?

This practice consists of selecting and designing effective summative assessments that allow the educator to gather valuable information about student learning. These assessments provide teachers and students with information that allows them to identify what students have learned, design strategies for planning future instruction, and make decisions to improve learning.

An evaluation has a summative intention when the educator uses evidence of learning to account for the achievement of students during a period of time (unit, semester, year). Sufficient evidence must be collected to determine the level of achievement of each of the defined learning processes and to certify them (Sanmartí, 2007).

Selecting and designing formal assessments of student learning involves (Förster, 2018):

1. Identify the learning objectives or expected learning of the students. It also involves constructing and/or selecting evaluation indicators that make explicit in evident terms the action(s) that a student must perform in order to consider that he/she has achieved a learning objective.

2. Designing the evaluation strategy and defining the necessary evidence. It involves deciding how best to evaluate content, skills and attitudes before teaching them. An evaluation strategy considers:

- Determine who the assessment agents will be: the teacher (hetero assessment), the student him/herself (self-assessment) or peers at the same hierarchical level in the classroom (peer assessment). An evaluation may include more than one agent.

- Select the evaluation procedure. Among the procedures or also called evaluation situations are: portfolio, case studies, projects, graphic organizers, research, essays, reports, etc. These procedures allow the evaluation of conceptual, procedural and metacognitive contents, as well as higher order skills.

- Select and design instruments. Evaluation instruments include:
  - Test-type situations: closed, open-ended (brief or extended) and mixed. While the first allows for the evaluation of factual, conceptual and, in some cases, procedural knowledge, open-ended questions allow for the evaluation of higher order skills (analyzing, evaluating and creating).
  - Task performance record: assessment scales, checklists, rubrics, observation records. These instruments are at the service of the performance situations, as they allow their correction and qualification.

- Selecting a referent. The referent of an assessment refers to the element that will be used to make a value judgment about the level of learning obtained by a student. Among the referents are:
    ● Ideographic: the evidence is compared with the student's previous performance.
    ● Normative: the evidence is compared with the performance obtained by other students who were evaluated (e.g., by students of the course or generation).
    ● Criterial: the evidence collected during the evaluation of students is compared with the evaluation objectives and indicators.

When constructing summative assessments, the educator must safeguard the following quality criteria:

Validity (Förster & Rojas-Barahona, 2008):
    ● Content validity: the instrument should collect information on what it is intended to assess (knowledge, skills and/or attitudes).
    ● Instructional validity: the instrument should include questions and activities that students have had the opportunity to learn during the classes. It is directly related to content validity.
    ● Consequential validity: the consequences derived from the application of an evaluation instrument must be in accordance with the purpose for which it was created.

Reliability: The assessment must be accurate. To this end, the evidence collected with an evaluation instrument must be sufficient to make a judgment on the level of achievement of their learning. Likewise, care must be taken with the conditions of application of the instrument, avoiding aspects that may deconcentrate students (such as noise, poor lighting, poor ventilation, etc.), ensuring that they have enough time and the quality of the instrument (handwriting, quality of the images, etc.), so that there are no external factors that affect their performance (Förster & Rojas-Barahona, 2008).

Objectivity: Both the evaluation instruments and the judgment made on the basis of the information gathered with them must be impartial and free of bias.

What is not Selecting and designing formal evaluations

➢ Make evaluations to regulate student behavior. Assessment should be used to make decisions to improve learning and not to control the classroom climate.

➢ Designing assessments that constitute an isolated milestone and that do not relate to what students had the opportunity to learn, practice and exercise during the course of the didactic unit or other teaching period. Failure to ensure coherence between the learning process and assessment runs the risk of students not having the opportunity to demonstrate how much they have learned. Assessment should be an integral part of the learning process (Förster, 2018). Therefore, planning must take into account in an integrated way both the design of teaching and the design of assessment strategies.

➢ Design only one assessment or question to assess a learning. More than one assessment method and/or question should be used within a test to ensure obtaining a more complete picture of a student's mastery of a set of learnings (Förster, 2018). This allows for increased reliability, having consistent trends about the student's performance and reducing inconsistencies caused by measurement errors that are caused by external factors, such as momentary emotional instability, illness, etc.

➢ Selecting the assessment without considering the skills, content and/or attitudes that should be mastered by the students in the unit to be assessed. By not considering the learning objectives, there is a risk of not effectively assessing what should be assessed and that the results will not be useful for decision making.

➢ Always use the same type of assessment procedure and/or instrument. Assessment influences the way in which students learn (Sanmartí, 2007); therefore, a variety of instruments and challenges that are relevant to the age, developmental stage of the students and the content (declarative, procedural and/or attitudinal) to be assessed should be presented. Otherwise, students will condition their learning to the type of assessment used by the teacher.

➢ Design evaluations in which heteroevaluation prevails. Students should be participants in their evaluation process, for which peer and self-evaluation should be used; this will promote reflection exercises, metacognition and will allow them to know themselves (Anijovich, 2012). Likewise, the participation of students will allow them to obtain more sources of information to make the value judgment related to the evaluation.

➢ Use assessments designed for other students or other contexts. If assessment instruments made in another context are translated or adapted, they must be adjusted to the reality and characteristics of the students. Otherwise, errors and distortions can be generated in the results as a result of misunderstanding (Förster, 2018).

Teaching strategies for selecting and designing formal assessments

● **Define purposes:** define the purposes of the evaluation by clearly identifying what decisions will be made based on its results. For example: adjust the planning, carry out reinforcement for some students, provide more in-depth activities, etc.

● **Determine the referent:** define how the results will be interpreted: criterial, normative or ideographic referent, according to the evaluation objectives: to evaluate the students' learning in relation to themselves (ideographic), in terms of the course performance (normative) or in terms of previously established criteria (criterial). For example, when using a criterion referent to evaluate a research paper, the teacher will compare each student's performance with a set of criteria (contained in a rubric or assessment scale) that arise from previously defined evaluation indicators. If you want to use an ideographic reference, you should compare the student's performance using two submissions of the research, for example, a draft that receives feedback and a final product that integrates the observations made.

● **Determine the agent:** determine whether the evaluation will involve the teacher, the students or both. The decision must be made thinking about the type of information to be collected and weighing which agent can provide it. To do this, it should be clearly determined which criteria will be used to evaluate each of them. For example, in the group elaboration of a model of the digestive system, students can evaluate the participation and responsibility of each of the members (co-evaluation), while the teacher can do it based on aspects such as the incorporation and representation of the structures of the digestive system (heteroevaluation). Finally, each student can evaluate if he/she fulfilled the role assigned to him/her within the group and if he/she actively participated by giving ideas and/or solving problems (self-evaluation).

- **Determine the assessment procedure and instrument:** in order to select the most appropriate ones, the nature of the content, skills and attitudes to be assessed should be taken into consideration (Armstrong, 2010). It is suggested to use a learning taxonomy that allows for a clear definition of the types of knowledge and skills to be assessed. For example, to evaluate the ability to "argue", it is suggested that students be presented with a situation or problem from which they must choose a position and support it based on the contents reviewed during the course of the classes.

- **Design evaluation strategies that consider authentic contexts:** This involves evaluating knowledge, procedures and/or attitudes using real problems or situations, so as to give value to knowledge for reflection and action (Diaz Barriga, 2005). For example, to evaluate an objective that involves performing calculations involving the four operations, propose to students that they make a budget for the end-of-course trip, arguing different options and the monthly fee that should be paid to finance it.

- **Review exemplars:** look for examples of evaluations of the selected objectives and indicators. To do this, you can review the programs of the Ministry of Education or other examples found online. It is always suggested to use validated instruments from recognized institutions.

- **Select standardized instruments:** A standardized assessment instrument has norms for the creation of the test, its application and the interpretation of its results. Since they have been made through a rigorous process that includes the participation of experts and their piloting, they have high rates of validity and reliability. They also allow for comparison of results between different samples of students (Backhoff, 2018). To select the most suitable standardized assessment, one must review the construct (content, skill and/or attitudes) that measures its purposes and have complete clarity on how to interpret its results.

- **Design evaluations that consider the following elements:**

  - **Ensure the coverage of the assessment:** for test-type situations, prepare a table of specifications of the instrument specifying the objectives, skills and contents of each item as well as the expected response (Förster & Rojas-Barahona, 2008). Review in contrast with the objectives of the unit to avoid leaving out a relevant aspect.

  - **Ensure rigor:** ensure that the assessment is challenging, including questions and activities that allow the application, analysis and reflection of the contents according to the students' abilities. Likewise, they should be aligned to the curriculum and to national and international standardized assessments (Bambrick-Santoyo, 2018). It is suggested to reflect together with other teachers to adjust the rigor of the evaluations of all subjects.

  - **Create evaluation guidelines and rubrics:** create, prior to the application of the instrument, guidelines and/or rubrics with a vocabulary understandable by the students and socialize them so that they can develop their work in a guided manner. All aspects that will be evaluated and graded should be included. For example: in a written essay, writing, use of evidence, references, format aspects, etc. will be evaluated. In the case of a practical performance, such as conducting a laboratory, in addition to evaluating the hypothesis statement, experimental design and discussion of results, procedural aspects such as the use of equipment and safety measures within the laboratory should be evaluated.

  - **Create guidelines and rubrics for self-evaluation or peer evaluation:** they should be clear, easy to apply and with concrete and observable criteria. They should also be clear for the age of the students. For example, in younger grades, graphic scales or scales containing symbols can be

used instead of numerical scales. It is also suggested to use economy of language, setting out guidelines with concise criteria.

- **Adjust the assessment:** adjust the instruments to the background and previous experiences of the students. Avoid biases caused by external factors to the assessment such as culture, developmental stage, ethnicity, gender, socioeconomic status, language, special interests, and special educational needs (Förster, 2018). To this end, situations or contexts, examples, images, and vocabulary used, among other things, should be reviewed.

- **Ensure instructional validity:** include questions or activities that are similar to the activities performed by students in class to avoid putting them in a disadvantageous situation (Förster & Rojas-Barahona, 2008). For example, in the subject of mathematics, problems that students have worked on during class can be adapted by changing the data or the context in which they were presented.

- **Ensure semantic validity:** use a language known and worked on during the unit or classes by the students to avoid errors due to misunderstanding of instructions or questions. It is suggested to have a record of the concepts that were used during the classes, as a glossary, for new words that may confuse the students. Likewise, in cases or problems that students must solve, the teacher should select concepts of everyday use (for example, use the word pencil instead of pen) or add a footnote with a definition if necessary.

- **Ensure content validity:** submit the evaluation to "judges" criteria. Ask the technical director or another teacher to review the evaluation instrument and, if necessary, make changes based on his or her suggestions.

- **Ensuring reliability:** ensuring that the environment in which an assessment is administered is free of distractions and pressures that interfere with student performance.

- **Increase reliability:** assess students repeatedly on the same learning, either in different instances or on several items within an assessment situation.

- **Ensure accuracy of correction:** assign, prior to the application of the instrument, the scores according to the relevance and level of complexity of the learning.

- **Ensure objectivity:** develop response or correction guidelines prior to the application of the instrument to ensure an objective reference at the time of correction.

- **Reduce bias:** in the case of written responses, it is advisable to follow an order in the correction that allows the performance of all students to be evaluated with the same rigor. For example, review all the answers to question 1, then continue with question 2 and so on.

References

Amstrong, P. (2010). Bloom´s Taxonomy. Vanderbilt University Center for Teaching. Recuperado el 28 de octubre de 2022 de https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy/

Anijovich, R. (2012). Evaluar para aprender. Conceptos e instrumentos. Aique Educación.

Backhoff Escudero, Eduardo (2018). Evaluación estandarizada del logro educativo: contribuciones y retos.Revista Digital Universitaria (RDU). Vol. 19, núm. 6 noviembre-diciembre. DOI: http://doi.org/10.22201/codeic.16076079e.2018.v19n6.a3

Bambrick – Santoyo, P. (2018). *Las Palancas del Liderazgo Escolar 2.0. Una guía práctica para construir colegios exitosos*. Corporación de la Educación Aptus Chile.

Bowen, R. S. (2017). Understanding by Design. Vanderbilt University Center for Teaching. Recuperado el 26 de octubre de 2022 de https://cft.vanderbilt.edu/understanding-by-design/.

Díaz Barriga, Frida. (2005). Enseñanza situada: Vínculo entre la escuela y la vida. México: McGraw Hill.

Fisher, M. R., Jr., & Bandy, J. (2019). Assessing Student Learning. Vanderbilt University Center for Teaching. Recuperado el 26 de octubre de 2022 de https://cft.vanderbilt.edu/assessing-student-learning/ .

Förster, C. (2018). *El poder de la evaluación en el aula. mejores decisiones para promover aprendizajes.* Ediciones Universidad Católica de Chile.

Förster, C. & Rojas-Barahona, C. (2008). Evaluación al interior del aula: Una mirada desde la validez, confiabilidad y objetividad. *Rev. Pensamiento Educativo*, Vol. 43, 2008. pp. 285-305.

Ministerio de Educación de Chile (2018). *Orientaciones para la implementación del Decreto 67/2018 de Evaluación, Calificación y Promoción Escolar. https://bibliotecadigital.mineduc.cl/bitstream/handle/20.500.12365/14279/orientaciones%20decreto%2067.pdf?sequence=1&isAllowed=y*

Sanmartí, N. (2007). 10 ideas clave. Evaluar para aprender. Graó.

Yale Poorvu Center for Teaching and Learning (s.f). *Formative and Summative Assessment.* Recuperado el 26 de octubre de 2022 de *https://poorvucenter.yale.edu/Formative-Summative-Assessments*

Suggested citation for this document

Observatorio de Buenas Prácticas Pedagógicas, Facultad de Educación, Universidad del Desarrollo (2023). "Selecting and designing formal evaluations".

This document was prepared by Antonia Ortega and reviewed by Kiomi Matsumoto, María Luisa Salazar and Catalina Correa within the framework of the Observatorio de Buenas Prácticas Pedagógicas, Facultad de Educación, Universidad del Desarrollo, Santiago, Chile.

Translated by María José Galmez, July 2023