



Initial Diagnostic Assessment in Pre-Service Teacher Training in Chile and its Relationship With Institutional Contexts

Evaluación diagnóstica inicial en la formación inicial docente en Chile y su relación con contextos institucionales

Valentina Giaconi^{1,2}, Gabriela Gómez¹,
Daniela Jiménez², Benjamín Gareca¹, Francisco Durán del Fierro³,
& María Leonor Varas²

¹ Universidad de O'Higgins

² Universidad de Chile

³ University College London

Abstract

This paper examines the way in which five Chilean universities have implemented the mandatory initial diagnostic assessment in pre-service teacher training. The law states that the outcomes of this assessment will be used to enhance students' learning processes. In addition, universities are entitled to autonomy to define what to evaluate and how to carry it out. Considering the purpose and scope of this evaluation, the aim of this paper is to better understand how these assessments have been implemented by universities and how institutional settings have affected the way in which universities have put this assessment into practice. In order to explore this phenomenon, the research employed a multiple case study approach. In this context, semi-structured interviews were carried out with institutional representatives, along with analysis of a set of diagnostic assessments and their psychometric test results. Drawing upon the triangulation of the information generated, the paper concludes that the most important elements affecting the implementation of this initial diagnostic assessment are, on the one hand, professionals' technical skills and, on the other, the resources that teacher training programs or faculties have. There was a lack of relationship between the assessment and the admission profiles and training programs.

Keywords: diagnostic assessment, teacher training programs, case study

Post to:

Valentina Giaconi
Avenida Libertador Bernardo O'Higgins 611, Rancagua, Chile
valentina.giaconi@uoh.cl
56 9 5135 2157
ORCID: 0000-0002-5166-5673

© 2022 PEL, <http://www.pensamientoeducativo.org> - <http://www.pel.cl>

ISSN:0719-0409 DDI:203.262, Santiago, Chile doi: 10.7764/PEL.59.1.2022.4

Resumen

En este artículo se estudia la implementación, en cinco universidades chilenas, de una ley que mandata la aplicación de una evaluación diagnóstica inicial en las carreras de pedagogía. La ley indica que los resultados de esta evaluación serán referenciales y tendrán un uso formativo para los estudiantes; asimismo, da autonomía a las universidades para definir qué y cómo evaluar. Considerando la importancia de esta evaluación y su alcance, el objetivo de este artículo es entender cómo se implementan estas evaluaciones y cómo su aplicación se relaciona con el contexto de cada institución. La metodología utilizada corresponde al estudio de casos y los datos analizados corresponden a entrevistas a representantes institucionales, al análisis de las evaluaciones efectuadas y a sus resultados psicométricos. Se concluye que los aspectos contextuales más importantes para definir las evaluaciones diagnósticas fueron las capacidades técnicas y los recursos a los que pudieron acceder las escuelas de pedagogía y se observó una desarticulación con los perfiles de ingreso o programas formativos.

Palabras clave: evaluación diagnóstica, formación inicial docente, estudio de caso

Current Chilean law and diagnostic assessment in teacher training programs

This paper studies the implementation of a policy in Chile that legally obliges all teacher training institutions to conduct diagnostic assessments at the start of their programs. The law stipulates that the results of these evaluations will be referential and will have a formative use for students (Ley N° 20.903). Formative assessment in initial teacher training is of key importance for various reasons. First, it is an essential tool to promote learning among student teachers (Brookhart, 2017). Second, modeling evaluation practices of this nature in initial teacher training is crucial to having future teachers who use this practice appropriately (Brookhart, 2017). Third, its use allows the identification of the key aspects that need to be monitored during training to ensure achievement of learning, just as in summative assessment (Klassen et al., 2017). However, little is known about the implementation of this type of initial diagnostic assessment with a formative focus.

Implementation of the policy studied in this paper began in 2017. If a university does not apply the assessment, it fails to comply with one of the regulations established in the mandatory accreditation of teacher training programs, which could result in its closure. This policy is novel, in the sense that assessment at the start of initial training is normally used for the purpose of filtering and selection, rather with a formative objective (Klassen et al., 2017; Sato & Kemper, 2017; Klassen & Kim, 2018). In Chile, students entering initial teacher training have already gone through a national admissions system that is used by most teacher training programs and which includes the application of exams based on the school curriculum (Santelices et al., 2019).

Furthermore, assessment generally has to comply with measurement standards and be supported by evidence of validity and reliability (American Educational Research Association et al., 2018). In the case of diagnostic assessment, Law N° 20,903 does not consider whether teacher training institutions have the necessary teams, capacities, and resources to adequately implement these assessments. It is therefore possible that the institutions face difficulties to implement the evaluations adequately and that they may not reflect the spirit of the law. Evidence of this should be provided in the texts and discourses associated with the evaluation systems of each institution (Daugherty & Ecclestone, 2006).

In order to comply with quality standards in measurement and have relevant evaluations, the role of institutions is key, where different levels (students, academics and teachers, administrative staff, and the labor market) must participate and be coordinated (Banta & Palomba, 2014). Evaluations are also affected by external influences (e.g., national and international policies) in addition to the internal institutional context (Kohoutek, 2014; Raaper, 2017; Daugherty & Ecclestone, 2006). Characterizing the effects of these contexts is essential in order to gain a better understanding of the implementation of these evaluations, which are embedded in a complex system (Flórez Petour, 2015; Daugherty & Ecclestone, 2006). Specifically for these assessments, the relevant law gives the institutions a great deal of autonomy to design and implement them, which provides an opportunity to analyze how the relationship between assessments and institutions has been established. Considering the importance of formative assessment in initial teacher training (ITT), and the originality and scope of the educational policy that obliges Chilean institutions to carry out diagnostic assessments of teacher training, a study was carried out, on which this paper is based, in which the following research questions are considered:

1. How have diagnostic assessments of teacher training been developed?
2. How does the institutional context affect the implementation of diagnostic assessments?

Evaluation and initial teacher training in Chile

Since the 1990s, various efforts have been made to improve the quality of initial teacher training in Chile (Avalos, 2014). For example, between 1997 and 2002, the Program for Strengthening Initial Teacher Education (ITT) was implemented, which provided competitive funding to institutions that wanted to improve their programs. However, during the 2000s, ITT was subject to an expansion that was largely deregulated, which resulted in the creation of many new programs, some of which provided unsatisfactory training processes (Ávalos, 2014). In the last 15 years, as a result of social demands, the need has arisen to drastically improve education in Chile and, accordingly, the quality of teaching (Avalos, 2014). This has highlighted the need for better teachers, which entails implementing policies to strengthen ITT based on their effectiveness according to comparative studies, which have mainly been done in three areas (Ingvarson & Rowley, 2017). The first of these involves increasing the selection of students with high academic performance, which is done by raising the minimum requirements for admission to teaching programs. The second area is related to the monitoring and establishment of standards for the implementation of ITT programs, achieved through mandatory accreditation and the reform of new requirements and standards. Lastly, the third area involves the evaluation and certification of students that are close to graduation, which is the reason why the National Diagnostic Assessment of Initial Teacher Training (ENDFID by the Spanish acronym) was implemented, which is applied shortly before student teachers graduate (Centro de Perfeccionamiento, Experimentación e Investigaciones Pedagógicas, 2018), but does not enable them to practice the profession. The initial diagnostic assessment on which we focus involves the first and second points, since (1) it allows steps to be taken regarding students who may not be sufficiently prepared for admission and (2) it enables improvement to the quality of ITT programs, since diagnostic assessments provide information that can allow curricula and teaching strategies to be adjusted.

Finally, the initial diagnostic assessment is indirectly linked to two other assessments that are compulsorily applied to the same population of students. On the one hand, in the university admission process, immediately prior to admission, they must have taken the selection tests (Santelices et al., 2019), the results of which could show gaps in areas of knowledge and skills, so a diagnostic assessment applied a few months later will only be useful if it adds information to that already collected. On the other hand, the formative objectives

of teacher training courses are outlined in nationally established standards, based on which the ENDFID evaluation is carried out, which is applied to students who are close to graduation (Centro de Perfeccionamiento, Experimentación e Investigaciones Pedagógicas, 2018).

The university selection tests and the ENDFID are standardized, while the initial diagnostic assessment is completely open and the universities have autonomy on how to implement it. However, this creates risks, because it is not certain that the universities have the necessary capabilities and teams to implement the evaluation adequately. Yorke (1998) states that in order to have effective management of evaluation in higher education, three conditions have to be met: (1) a clear definition of the purpose of evaluation, (2) a strategy to achieve this purpose, and (3) implementation that works. This is also outlined by the Standards for Educational and Psychological Testing (American Educational Research Association et al., 2018), which define validity as “the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests” (p. 11), meaning that if a test lacks a defined purpose, it is not possible to make a validity judgment. In this respect, it is essential to identify clear and bounded purposes. Furthermore, the initial diagnostic assessment is defined with a formative purpose, so, for these evaluations to be valid, they must promote learning and be associated with certain actions (Heritage et al., 2009). Therefore, the associated validity argument has to consider the consequences of these evaluations (Stobart, 2006). Likewise, the interpretation of the reliability of these evaluations has to consider their training focus. Black and Wiliam (2006) argue that reliability is a less relevant aspect in formative assessments if they are formed by frequent interactions between students and teachers; however, in the case of a formative test that will allow decisions to be made about how much time a student has to spend on a remedial course, reliability becomes a crucial aspect. There is evidence of successful formative evaluations in university contexts, such as experiences with writing assessments at the beginning of courses (Fisher et al., 2011), formative evaluations applied in teacher training courses (Cosi et al., 2020), and interventions associated with formative evaluation, with different characteristics regarding the involvement of teachers and students and the use or not of technology (Stull et al., 2011).

In the case of the initial diagnostic assessment for teaching training, the law outlines formative and, more specifically, remedial uses related to standardization of prior knowledge. Besides these uses, Law N° 20,903 does not state other purposes, nor does it specify the necessary knowledge and skills to be assessed, so each university has to define its own initial evaluations. However, because of this, it is not clear whether universities have the necessary resources, nor that they are able to implement evaluations with sufficient evidence of validity and reliability. It is therefore necessary to assess whether the formative uses of these evaluations are defined and if they show evidence of validity and reliability.

Finally, it is important to underline that this policy of initial diagnostic assessment is unique in the international context, where evaluations for teachers tend to be for the purposes of selection and certification rather than diagnosis (Klassen et al., 2017; Sato & Kemper, 2017; Klassen & Kim, 2018).

In line with this, it is essential to understand how these initial diagnostic assessments have been developed for teacher training courses.

Assessment and its relationship with external and institutional contexts

It is essential to take into account the external and institutional context in order to understand how diagnostic assessment functions and the differences in these assessments between the institutions (Kohoutek, 2014; Biggs 1996). Kohoutek (2014) suggests a conceptual framework with four elements in order to understand student assessment in higher education, which includes institutional policies and regulations, academic management and organization, verification strategies (e.g., dual review of assessments), and external influences on assessment,

which include national regulations and international laws. Raaper (2017) studies two European universities and shows how aspects that are specific to the institutions define their evaluative policies, such as age, traditionalism, and internal policies, as well as national and global contexts related to neoliberalism. On the other hand, Flórez Petour (2015) shows how structures that are more extensive than the institutions themselves (such as political parties, the economic sector, families, and the media) also determine the reforms associated with formative assessment in Chile. These studies show the importance of studying the implementation of evaluations, considering the political and institutional context.

In the case of the initial diagnostic assessment of student teachers, the institutions have to adapt to the law that makes the implementation of these evaluations obligatory, taking into account their internal policies and organization (Ley N° 20.903). Biggs (1996) contends that an institution and its educational practices have to be balanced in order to achieve its educational objectives. Indeed, for an evaluation to be used effectively to improve student learning, it must include the participation of academics, administrative staff, students, and employers (Banta & Palomba, 2014; Richmond et al., 2019). Banta and Palomba (2014) define the roles of each of these groups, as described below. They consider that academics and students have an essential and pivotal role, and should be heavily involved in evaluation. Academics have to be responsible for carrying out all stages of the evaluation process, from defining the objectives to taking action on the results. The main role of students is to participate in evaluations and act on their results, but it is also important for them to participate in committees and give feedback on evaluations and teaching. In order to encourage participation, the role of administrative staff is necessary to support academics and students, promote the use of the results in development of the curriculum, bring in knowledge (e.g., experts, promote attendance at conferences), and provide funding. Meanwhile, employers can provide important perspectives on what should be assessed.

In short, the role of the general (global, country-level) and institutional contexts are essential to understand the implementation of diagnostic assessments. Specifically, it is necessary to obtain information on the relevant actors involved in the evaluations, such as administrative staff and teachers, as well as students, whose main responsibility is to respond to the assessments.

Methodology

This study uses a qualitative methodology that consists of multiple cases. Each case corresponds to a diagnostic assessment process for teacher training, implemented in a university that provides initial teacher training programs.

This methodological approach is appropriate because (1) the focus of the research is to understand how these evaluations are implemented, as they are a complex process that depends on the institutional context, and (2) there are various sources of information for each case (interviews, documents, evaluations, and psychometric results). Study of multiple cases allows us to understand each evaluation process in its institutional context (Yin, 2011), but also enables us to produce greater generalization of the findings, since the objective is to observe common processes and results among the cases, understanding what is conditioned by local realities (Miles et al., 2013).

Sample

The sample consists of five universities that cover the majority of the enrollment in teacher training courses in Chile. The universities are diverse, since they are both public and private (two are public and three are private), they have different geographical locations (two in the Metropolitan Region and three in other regions), have different numbers of years of institutional accreditation (ranging from no accreditation to the maximum number of years of accreditation), they have a variety of programs, and not all are members of the

Council of Rectors of Chilean Universities (CRUCH) (four are and one is not). The sample was selected in such a way as to have a diversity of institutions in accordance with the aforementioned characteristics, and universities with high levels of enrollment and a variety of pedagogy programs were also considered to ensure the psychometric analysis of the instruments.

Data collection

The data were collected in 2018 and include information on the diagnostic assessment process implemented in 2017.

The institutions chosen were contacted and invited to participate in the study. The rector of each institution authorized its participation in the study. After reaching agreement, the following was requested from each institution:

- Interview three institutional profiles with the objective of understanding the development and implementation of diagnostic assessments. The profiles correspond to (1) a managerial position, responsible for making decisions regarding the assessments, (2) a technical-operational position responsible for implementation of the assessments, and (3) a teacher who could potentially use the results of the assessments. The interviews were semi-structured and included questions about the evaluation process (frame of reference of the diagnostic assessment, characteristics of the instruments, and operational aspects) and the institution (students and their interests and requirements).
- The instruments used for the diagnostic assessment.
- All available documentation associated with the assessment (frameworks for evaluation, application protocols, rubrics, etc.).
- The databases with the results of the instruments applied.

The data collection processes followed strict security and data protection protocols. All interviewees agreed to participate in the research by signing an informed consent form.

Analysis

In order to study the cases, all sources of information (interviews, instruments, etc.) were considered, which made it possible to describe the development of each evaluation system depending on its context.

For each case, the interviews were analyzed following the guidelines of the thematic analysis technique. This was decided because this analytical approach allows the production and interpretation of central themes and the analysis of themes rather than individual experiences (Joffe, 2011). In this respect, a system of codes was created, the analysis categories of which were defined inductively based on reading a subset of the interviews.

In order to validate the codes, the first case was coded independently by two members of the research team. These codes were compared and the definition of the codes was refined according to the differences found. With the redefined codes, the first case was analyzed again, obtaining full agreement between the researchers, so the coding system could be considered validated. All of the interviews were then coded and reports were subsequently produced for each case. The texts produced thus allowed the reconstruction of the contexts and themes associated with the diagnostic assessments, which contributed to answering the research questions.

The analysis of the cases also included review of a set of materials developed by the institutions aimed at supporting the processes to design and apply the tests. A content analysis was also done on the language, mathematics, and non-disciplinary aspects of the assessments through a review by a panel of experts (Sireci & Faulkner-Bond, 2014). These areas were examined by a feasibility criterion and because language, mathematics, and non-disciplinary aspects were areas assessed at all institutions. A panel of experts was formed for each area,

consisting of mathematics teachers, language teachers, and educational psychologists, respectively. This panel assessed the conceptual relevance of the instruments and their items in relation to the discipline being evaluated and their intended use. Similarly, the panel assessed measurement aspects (clarity of the items, number of items per dimension, etc.). In order to do this, an item and instrument review protocol was used and psychometric analyses were carried out of the mathematics and language instruments. These analyses were based on classical test theory (Martínez-Arias, 1996), and Cronbach's alpha coefficient was used to estimate the reliability. The details of these analyses can be found in Giaconi et al. (2019).

Results

Table 1 shows a summary of the main characteristics of each case according to the dimensions of analysis. This section uses the results in Table 1 to describe the cases with regard to the dimensions of analysis that are clearly related to the evaluative processes, and then provides a general analysis of the dimensions that did not show an explicit relationship with the evaluations. The first case presented (Case 2) has the most developed evaluative system and the last case (Case 5) the least developed.

Case 2

This case showed the greatest development in its evaluation system. The development of diagnostic assessments was carried out in line with a decision by the institution that was taken long before the implementation of Law 20,903. This university is a highly complex institution, which has had accreditation for many years. Most of its instruments were applied to the entire university, and the interviewees underlined that the articulation between different institutional units with diverse technical capacities (for example, in evaluation and technological developments) was essential for the creation of a good assessment system. The university also had explicitly defined purposes and uses for its evaluations. It is possible that these factors explain why it was the only institution that produced good psychometric indicators for its assessments and for which the panel of experts considered the content of the evaluations to be adequate.

Case 3

This case showed that a lot of work had been done to implement specific evaluations for its teaching courses. It was the only case in which an entrance profile had been defined for these courses that provided support for the process of creating instruments. This institution also managed to implement an online application. This case was also a complex university with several years of accreditation. The progress achieved is valuable, as it is in line with having a system that is very well adapted to the needs of the teacher training courses. However, as a whole, the case shows a medium level of development and requires improvement, since the evaluations could not be piloted previously, they displayed certain problems of construction, the psychometric indicators were inadequate, and the use of the evaluations was relatively unspecific. It was also reported that the workload exerted stress on a small team associated with teacher training courses.

Case 1

In this case, the institution constructed evaluations and adapted assessments from other projects. It also received technological support for the application and use of the evaluations. However, the objectives and uses are described with little specificity and the interviewees did not report any use for specific actions. Problems

were observed in the construction and psychometric properties of the instruments; for this reason, we consider that the level of development was also medium. There is no articulation with the training model, nor did we observe the existence of a review process for these assessments.

Case 4

This case, like case 3, stands out due to the institution having made an effort to create a specific system adapted to its teacher training courses. It also had systematized processes and protocols for tracking and recording processes. However, there were significant flaws in the design of the evaluations and in the psychometric indicators, so we consider that it showed medium development. We also observed somewhat general uses, which did not have explicit connections with the assessments.

Case 5

This case showed the lowest level of development, as most of the process was outsourced and performed by another institution. It also applied the same tests throughout the university. This was justified because the institution had not yet developed the capabilities to design and implement its own diagnostic assessment system. We also observed that the uses were poorly specified and, in fact, there was no articulation between the remedial work done by the university and the results of the diagnostic assessment.

In the results we can see that assessment systems need further development in most cases, particularly regarding (1) the definition of clear purposes and uses that are connected with the design of the assessments, and (2) the improvement of psychometric and design characteristics. This enables us to infer that there is a lack of capacity and resources at universities to develop better assessments and systems for application and use. They also need time to evaluate and improve assessment systems

Finally, it is important to note that no relationship or articulation was observed between dimension 1 (institutional policy and training model) and dimension 2 (student profile) with the diagnostic assessment systems. These issues were mentioned by the interviewees, but were not used when describing the design processes or the use of the evaluations.

Table 1
Summary of results of cases

Dimension	Case 1	Case 2	Case 3	Case 4	Case 5
1. Institutional policy and training model	1. Socio-critical perspective associated with social teacher training aimed at social transformation.	1. Theoretical-disciplinary training based on pedagogical knowledge, didactics of the specialty, and disciplinary knowledge. Progressive pedagogical practices.	1. Competency-based model.	1. Practice as the central pillar of learning, as the articulating aspect of the teaching process and knowledge production.	1. Modelo formativo común para las carreras de pedagogía, definido por: énfasis en formación disciplinar, vínculo con la práctica, dominio de las tecnologías de la información y de las comunicaciones (TIC), atención a la diversidad y reconocimiento del trabajo colaborativo.
2. Student profile	2. Homogeneous profile, since most of the students are from subsidized and municipal educational establishments. However, the distribution of scores in the university selection test is not homogeneous.	2. Profile oriented towards students from municipal and subsidized private schools and families in the first three income quintiles. Students come predominantly from the same region.	2. Profile with high performance on university selection test (PSU), coming from the same region and from subsidized private schools.	2. The student profile has changed as a consequence of the university's incorporation into the Single Admissions System, increasing the average scores of students and the number of students enrolling.	2. Perfil marcado por estudiantes de la misma región, primera generación en la universidad y provenientes de escuelas municipales y particulares subvencionadas.
3. Experience in diagnostic assessments of first-year students prior to Law 20,903.	3. Interviewees mention previous experiences in the design of diagnostic assessments.	3. Diagnostic assessments have been applied since 2014. Pioneering institution in the development of evaluations and the use of their results. Established mechanisms for collaboration between different institutional bodies.	3. Experience in the application of diagnostic assessments in all courses to understand university dropout. For this reason, they were not considered for the diagnostic assessment in teacher training.	3. No previous experience is mentioned.	3. No mencionan experiencias previas.

Dimensión	Caso 1	Caso 2	Caso 3	Caso 4	Caso 5
4. Areas evaluated by instruments	Most of the instruments were applied throughout the whole university.	Most of the instruments were applied throughout the whole university.	Most of the instruments were specifically for teaching training. Only case that prepared an admission profile for I.T.T.	Most of the instruments were specifically for teaching training.	Los instrumentos eran aplicados en toda la universidad.
4.1	4.1 Reading Comprehension and Language and Communication, Science, Mathematics, and Information and Communication Technologies.	4.1 Reading Comprehension and Language and Communication, Writing, Mathematics, Biology, Physics, and Chemistry.	4.1 Reading Comprehension and Language and Communication, Mathematics, Biology, Physics, Chemistry, Music, and Physical Education.	4.1 Reading Comprehension and Language and Communication, Writing, History, English and Physical Education.	4.1 Matemática y Escritura.
4.2 Non-disciplinary areas	4.2 Learning styles.	4.2 Learning strategies.	4.2 Pedagogical skills.	4.2 Metacognitive strategies and disposition or vocation of teacher.	4.2 No reportaron instrumentos de áreas no disciplinares.
5. Purposes and uses	To obtain information on the competencies and skills of students received. To find out the areas in which students require support and thus plan and adapt classes in accordance with gaps identified.	To characterize students' degree of mastery of school content on graduation from the school system, in order to generate institutional support actions. Specific uses: design and implementation of remedial actions; creation of the profile of first-year students differentiated by course for teachers, and formation of the evaluation model for each course.	To assess the basic skills that a teacher needs to develop. These skills would be assessed in order to establish support mechanisms to enable their development. Other expected uses for diagnostic assessments were to inform training processes and to alert teachers to the need to consider results in teaching practice. There is no assistance or tutoring focused on academic support linked to diagnostic assessments. Workshops for induction into university life are available for all first-year students.	To create improvement plans for teacher training courses. Uses: to enhance improvement plans, to find out about the students' entry profile, to support curricular innovation and quality assurance processes, and to monitor students throughout the training process.	Contribuir a la construcción de un perfil académico de los nuevos estudiantes que permitiera diseñar políticas institucionales. La institución externa encargada del diseño de las evaluaciones diagnósticas utilizó instrumentos que ya había elaborado con anterioridad. Los entrevistados no hacen referencia a los propósitos originales de aquellos instrumentos. Los diagnósticos no son considerados en el diseño de nivelaciones debido a la extensión del periodo de revisión de resultados exigido por la instrucción externa.

Dimensión	Caso 1	Caso 2	Caso 3	Caso 4	Caso 5
6. Process of creating instruments and 7. Application of the evaluations	<p>6. The institution creates evaluations and adapts evaluations from previous projects.</p> <p>7. Online application. The institution provides technological support to support the evaluation process and which allows tracking of students.</p>	<p>6. The design and preparation of diagnostic assessments is based on instruments used in previous years. There is a process of redesigning each instrument every two years on an alternate basis, according to institutional guidelines. Pilot applications are carried out as a validation process.</p> <p>7. Online application. Different units are connected (application of evaluations, support, etc.). Institutionalized processes and preparation of support materials.</p>	<p>6. The process of preparation, application, analysis, and reporting of the diagnostic assessments was carried out internally in the teacher training schools by a team led by a professor with expertise in evaluation.</p> <p>A common admission profile was created for all teacher training courses, which established pedagogical and disciplinary skills as a conceptual basis for the creation of the evaluations.</p> <p>7. Online application.</p>	<p>6. The design, implementation, and evaluation of diagnostic assessments is the responsibility of a unit that operates in a more or less centralized manner. General tests have been adapted for the teaching training courses.</p> <p>The benchmarks for the tests are varied and complementary processes are carried out to define the contents: bibliographic review, adjustments of items from the bank of questions, and use of items from other evaluations.</p> <p>7. Application using pen and paper, where the faculty professors/teachers play a key role. The university has an application coordinator for each site, application protocols, and systematized processes.</p>	<p>6. The instruments are created by an external institution. The institutional directorate responsible for academic affairs organizes the implementation of diagnostic assessments for all first-year students.</p> <p>7. Application using pen and paper, which is the responsibility of the external institution in charge of evaluations. The decision to outsource this process is comprehensible, as the university does not yet have the necessary capabilities.</p>

Dimensión	Caso 1	Caso 2	Caso 3	Caso 4	Caso 5
8. Content and analysis and psychometric analysis	<p>The items were appropriate for the instruments; however, the disciplinary tests included very few items, which prevented it from achieving acceptable reliability (Cronbach's alpha between 0.2 and 0.49).</p>	<p>The instruments were correctly created and a reference to the PSU was observed in the structure and content of the items. The mathematics and language assessments showed good indices of reliability (higher than 0.7).</p>	<p>The language tests were adequate; however, there are construction problems in the mathematics test and the test on non-disciplinary aspects. For example, the latter contained too few items for each dimension. The psychometric results indicated that the language and mathematics tests had low levels of reliability (range between 0.10 and 0.45).</p>	<p>Construction problems were identified; for example, combining disciplines in the same assessment or using too few items in each construct assessed. The psychometric results showed low levels of reliability in the set of language and mathematics items (range between 0.09 and 0.33).</p>	<p>The instruments were generally constructed adequately. In the psychometric analysis, the assessment showed a low reliability of 0.52.</p>

Note: The information presented in this table comes from the analysis of interviews and associated documentation (dimensions 1 to 7) and from the content analysis and psychometric analysis of the instruments (dimension 8).

Source: Prepared by the authors.

Summary and Final Discussion

From the analysis of these results, we can point to the themes associated with the research questions initially posed. With respect to the first question, which refers to the development of diagnostic assessments of teacher training, we can see varied levels of development of the diagnostic assessment systems at each institution. Case 2 had an institutionally established and comprehensive system, into which the teacher training courses were incorporated. Its objective was to assess knowledge acquired at school. The system was robust and connected between the central units of the university. As the evaluations were institutional, they were not applied specifically for teaching courses. This was the only institution whose tests had reliability indices higher than 0.7. Meanwhile, Case 3 implemented an independent system that was applied specifically to teacher training courses based on its own framework of theoretical evaluation. The process was developed based on a team that had the technical knowledge to carry out the task, but was heavily overloaded. In Case 1, the assessments were also applied to most of the university, which would eventually allow institutional capabilities to be used; however, they were seen to be less advanced than in Case 2. In Case 4, the university had decided to adapt a previously existing system that was applied across the institution. This system had systematized processes and protocols for tracking and recording processes. However, there were significant flaws in the design of the evaluations and their psychometric indicators. Finally, the institution in Case 5 had decided to outsource the design, implementation, and analysis of the system. The assessment used was an instrument created for more general purposes, linked to the assessment of first-year students and not specifically for teacher training. This decision was related to the lack of internal bodies that were able to carry out these tasks, thus reducing their capacity to influence the process.

With regard to the second question, the cases show that the degree of development of the assessment systems and their quality varied from one institutional context to another and depend, first, on the capacities previously established in each institution, second, on the definition of the purposes and uses associated with the assessments, and, third, on the decision to apply general tests across the entire university or specific tests for teacher training.

In relation to the first point, these capacities are highly complex and take time to develop, since they entail having formal processes, a tracking system, consolidated and stable teams, and technical knowledge of assessment. For example, the only university whose evaluations showed high reliability indices had had a diagnostic system in place for several years and used it to continue evaluating teacher training courses (Case 2). The rest of the universities implemented new, adapted, or outsourced systems and none of them had a totally satisfactory system in technical terms.

As regards the second point, establishing the purposes and uses of assessment and the subsequent connection of the assessments with these purposes and uses can be highlighted as a key aspect. Case 2 was the also one in which this articulation was better developed, where the purposes were defined with greater detail and specificity (to discern the degree of mastery of school content), which allowed evaluation to be carried out in line with these objectives. In the other cases, no explicit articulation was observed and the purposes were not clearly defined.

Finally, and with respect to the third point, among the cases studied, we can discern two ways to approach the implementation of assessment: general tests used for a variety of courses or specific tests for teacher training courses. This decision was related to the institutional context, since applying general tests allows a university to use its entire structure and capabilities, while applying specific tests is usually the responsibility only of the schools of education, which tend to be formed by smaller teams. This could explain why institutions that developed assessments specifically for teacher training had problems in defining the purposes and conceptualizing the tests. One example is Case 4, which attempted to adapt its general assessment system to teacher training and its evaluations suffered from design problems. Another example is Case 3, where the school of pedagogy assumed responsibility for carrying out the process, at the cost of overloading a small team.

One last result to highlight in terms of the development of evaluation systems is that there is no articulation between these evaluations and the training models, nor is there a relationship with the student profiles. The interviewees did identify the institutional training models and the student profile, but these elements are not mentioned when they explain the diagnostic assessment systems.

Law N° 20,903 stipulates that there should be diagnostic assessments for teacher training courses and makes them compulsory. This means that the corresponding state agencies should develop policies and guidelines to facilitate and guide their implementation. Like the universities, state agencies are in the process of establishing the guidelines, which leaves several questions unanswered. For example, the law states that the National Accreditation Commission has to verify the development of these diagnostic assessments (Ley N° 20.903), but it is not clear to what extent this diagnostic assessment should be aligned with the training models, target problems such as retention and academic performance, or aim at development of the curriculum. These questions are key, as diagnostic assessments partly define what student teachers should already know before starting their training (Klassen et al., 2017). Similarly, within the framework of the professional teacher development system (which is the responsibility of the Pedagogical Improvement, Experimentation and Research Center [Centro de Perfeccionamiento, Experimentación e Investigaciones Pedagógicas, CPEIP]) there are unresolved questions regarding its possible coordination with diagnostic assessments later in training, such as the ENDFID. On the other hand, the organization and development of diagnostic assessments is the responsibility of each institution. As a consequence, it is also the institutions themselves that can independently define the purposes of their own systems. This is why certain tension is generated between this autonomy and the possible guidelines and purposes established in the professional teaching system.

The literature shows that assessing students properly is difficult and costly in terms of time and resources, and also depends on many actors within a community (Banta & Palomba, 2014; Richmond et al., 2019) and requires institutions to have the appropriate operational conditions and specific technical knowledge. Therefore, in addition to the actions that have to be carried out by universities, it is essential for the state to provide human and material resources to support this policy. Our analysis confirms what the literature has highlighted with regard to the challenges associated with the development of assessment systems, emphasizing the importance of universities having teams that are sustained over time, which will enable the implementation and continuous assessment of the evaluations carried out. Similarly, the universities need to define more specific purposes and articulate their assessments in order to meet these objectives. In this respect, it can be expected that an evaluation system will not operate satisfactorily in the first few years after it is implemented.

The results of this study shed light on how the tensions between the law, public agencies, and universities emerge in fledgling and unconsolidated assessment systems, which should be considered in the accountability processes that quality accreditation commissions require from universities. In this respect, it is necessary for the institutions to be able to evaluate and improve their assessment systems before making decisions. It is also essential for universities to receive the necessary technical and material support in order to achieve these objectives. For this reason, it is indispensable to promote partnerships and collaboration between institutions. Considering the complexity of assessment and the need to test assessments in large samples, it is essential to optimize resources and promote reflection on shared needs.

Finally, with respect to the limitations of this study, the first is that the data correspond to the first year of implementation of the law. It is important to look at this period because it sheds light on the initiation of the law; however, follow-up is necessary, as these assessment systems may have evolved. A second limitation is that the purposes and uses of the evaluations were not clearly defined in most of the institutions, which makes it difficult to ascertain the relevance of the instruments applied. Considering the key role played by the use of these assessments, we believe that future research should focus on this issue. Lastly, a third limitation is that

the study does not allow us to understand the role of the relevant public agencies, which is very important in order to examine the implementation of evaluative policies (Flórez Petour, 2015; Stobart, 2006). In this vein, it would be interesting for future research to include representatives of these public agencies among the informants.

Funding: The research project on which this paper is based was funded by the Ministry of Education of Chile, through the Fund for Research and Development in Education (Fondo de Investigación y Desarrollo en Educación) (project code: FON170009, 2017).

Acknowledgements: This research was supported by the Mathematical Modeling Center (Centro de Modelamiento Matemático, CMM), ACE210010 and FB210005, and CIAE FB0003, BASAL funding for centers of excellence from ANID-Chile.

The original paper was received on April 20, 2020
 The reviewed paper was received on March 8, 2021
 The paper was accepted on August 9, 2021

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2018). *Estándares para pruebas educativas y psicológicas*. American Educational Research Association.
- Ávalos, B. (2014). La formación inicial docente en Chile: Tensiones entre políticas de apoyo y control. *Estudios Pedagógicos (Valdivia)*, 40(Especial), 11–28. <https://doi.org/10.4067/S0718-07052014000200002>
- Banta, T. W., & Palomba, C. A. (2014). *Assessment Essentials: Planning, Implementing, and Improving Assessment in Higher Education*. John Wiley & Sons.
- Biggs, J. (1996). Assessing Learning Quality: reconciling institutional, staff and educational demands. *Assessment and evaluation in Higher Education*, 21(1), 5-16. <https://doi.org/10.1080/0260293960210101>
- Black, P., & Wiliam, D. (2006). The Reliability of Assessments. In J. Gardner (Ed.), *Assessment and Learning* (pp. 149-167). SAGE.
- Brookhart S. (2017). Formative Assessment in Teacher Education. In D. J. Clandini, & J. Husu (Eds.), *The SAGE Handbook of Research on Teacher Education* (Vol. 2) (pp. 927-943). SAGE.
- Centro de Perfeccionamiento, Experimentación e Investigaciones Pedagógicas. (2018). *Resultados Nacionales. Evaluación Nacional Diagnóstica de la Formación Inicial Docente 2018*. CPEIP.
- Cosi, A., Voltas, N., Lázaro-Cantabrana, J. L., Morales, P., Calvo, M., Molina, S., & Quiroga, M. Á. (2020). Formative assessment at university through digital technology tools. *Profesorado, revista de currículum y formación del profesorado*, 24(1), 164-183. <https://doi.org/10.30827/profesorado.v24i1.9314>
- Daugherty, R., & Ecclestone, K. (2006). Constructing Assessment for Learning in the UK Policy Environment. In J. R. Gardner (Ed.), *Assessment and Learning* (pp. 149-167). SAGE.
- Fisher, R., Cavanagh, J., & Bowles, A. (2011). Assisting transition to university: Using assessment as a formative learning tool. *Assessment & Evaluation in Higher Education*, 36(2), 225-237. <https://doi.org/10.1080/02602930903308241>
- Flórez Petour, M. T. (2015). Systems, ideologies and history: a three-dimensional absence in the study of assessment reform processes, *Assessment in Education: Principles, Policy & Practice*, 22(1), 3-26. <https://doi.org/10.1080/0969594X.2014.943153>

- Giaconi, V., Varas, M. L., Ravest, J., Martin, A., Gómez, G., Quepil, J. P., & Díaz, K. (2019). *Fortaleciendo la Formación Inicial Docente: experiencia universitaria en la implementación de la Evaluación Diagnóstica Inicial para Pedagogías. Informe FONIDE 170009*. Ministerio de Educación de Chile.
- Heritage, M., Kim, J., Vendlinski, T., & Herman, J. (2009). From Evidence to Action: A Seamless Process in Formative Assessment? *Educational Measurement: Issues and Practice*, 28(3), 24-31. <https://doi.org/10.1111/j.1745-3992.2009.00151.x>
- Ingvarson, L., & Rowley, G. (2017). Quality Assurance in Teacher Education and Outcomes: A Study of 17 Countries. *Educational Researcher*, 46(4), 177-193. <https://doi.org/10.3102/0013189X17711900>
- Joffe, H. (2011). Thematic Analysis. In D. Harper, & A. R. Thompson (Eds.), *Qualitative Research Methods in Mental Health and Psychotherapy* (pp. 209-223). John Wiley & Sons. <https://doi.org/10.1002/9781119973249.ch15>
- Klassen, R., Durksen, T., Patterson, F., & Rowett, E. (2017). Filtering functions of assessment for selection into initial teacher education programs. In D. J. Clandini, & J. Husu (Eds.), *The SAGE Handbook of Research on Teacher Education* (Vol 2) (pp. 893-909). SAGE.
- Klassen, R. M., & Kim, L. E. (2018). Selecting teachers and prospective teachers: A meta-analysis. *Educational Research Review*, 26, 32-51. <https://doi.org/10.1016/j.edurev.2018.12.003>
- Kohoutek, J. (2014). European standards for quality assurance and institutional practices of student assessment in the UK, the Netherlands and the Czech Republic. *Assessment & Evaluation in Higher Education*, 39(3), 310-325. <https://doi.org/10.1080/02602938.2013.830694>
- Ley N° 20.903. Crea el sistema de desarrollo profesional docente y modifica otras normas. (2016, 1 de abril). <https://www.leychile.cl/Navegar?idNorma=1087343&idParte=>
- Martínez-Arias, R. (1996). *Psicometría: Teoría de los tests psicológicos y educativos*. Síntesis.
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2013). *Qualitative Data Analysis: A Methods Sourcebook*. SAGE Publications.
- Richmond, G., Salazar, M. D. C., & Jones, N. (2019). Assessment and the Future of Teacher Education. *Journal of Teacher Education*, 70(2), 86-89. <https://doi.org/10.1177/0022487118824331>
- Raaper, R. (2017). Tracing assessment policy discourses in neoliberalised higher education settings. *Journal of Education Policy*, 32(3), 322-339. <https://doi.org/10.1080/02680939.2016.1257160>
- Santelices, M. V., Catalán, X., & Horn, C. (2019). University Admission Criteria in Chile. In M. V. Santelices, C. Horn, & X. Catalán (Eds.), *The quest for equity in Chile's higher education. Decades of Continued Efforts* (pp. 81-90). Lexington Books.
- Sato M., & Kemper S. (2017). Teacher Assessment from Pre-service through In-service Teaching. In D. J. Clandini, & J. Husu (Eds.), *The SAGE Handbook of Research on Teacher Education* (Vol 2) (pp. 944-962). SAGE.
- Sireci, S., & Faulkner-Bond, M. (2014). Validity Evidence Based on Test Content. *Psicothema*, 26(1), 100-107. <https://doi.org/10.7334/psicothema2013.256>
- Stull, J., Varnum, S. J., Ducette, J., & Schiller, J. (2011). The Many Faces of Formative Assessment. *International Journal of Teaching and Learning in Higher Education*, 23(1), 30-39. <https://www.isetl.org/ijtlhe/pdf/IJTLHE851.pdf>
- Stobart, G. (2006). The Validity of Formative Assessment. In J. Gardner (Ed.), *Assessment and Learning* (pp. 149-167). SAGE.
- Yin, R. K. (2011). *Case Study Research and Applications: Design and Methods* (6a Ed.). SAGE.
- Yorke, M. (1998). The Management of Assessment in Higher Education. *Assessment & Evaluation in Higher Education*, 23(2), 101-116. <https://doi.org/10.1080/0260293980230201>