Document généré le 18 juil. 2025 23:03

Mesure et évaluation en éducation

Narrative of a Practice for the Adaptation and Use of a Training tool for Evaluative Judgment Based on Concordance

Récit de pratique sur l'adaptation et sur l'utilisation d'un dispositif de formation au jugement évaluatif basé sur la concordance

Relato prático sobre a adaptação e utilização de um dispositivo de formação do juízo avaliativo baseado na concordância

Marie-France Deschênes, Eric Dionne et Michelle Dorion

Volume 46, numéro spécial, 2023

Translation Issue

URI : https://id.erudit.org/iderudit/1118374ar
DOI : https://doi.org/10.7202/1118374ar

Aller au sommaire du numéro

Éditeur(s)

ADMEE-Canada

ISSN

0823-3993 (imprimé)
2368-2000 (numérique)

Découvrir la revue

Citer cet article

Deschênes, M.-F., Dionne, E. & Dorion, M. (2023). Narrative of a Practice for the Adaptation and Use of a Training tool for Evaluative Judgment Based on Concordance. *Mesure et évaluation en éducation*, *46*(spécial), 1–21. https://doi.org/10.7202/1118374ar

Résumé de l'article

Les futurs enseignants ont rarement l'occasion de tester leurs interprétations de questions complexes, incertaines ou ambiguës en évaluation des apprentissages et d'obtenir une rétroaction formative sur ces questions. Dans cet article, nous présentons un récit de pratique sur l'adaptation et sur l'utilisation d'un dispositif de formation au jugement évaluatif de futurs enseignants basé sur la concordance de jugement. Les assises théoriques du dispositif, les principes de sa conception ainsi que les facteurs contextuels ayant influencé son adaptation sont abordés. Un tel exercice nous permet de cerner quelques avantages de l'utilisation du dispositif pour favoriser le développement du jugement évaluatif en éducation.

© Marie-France Deschênes, Eric Dionne et Michelle Dorion, 2025

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter en ligne.

https://apropos.erudit.org/fr/usagers/politique-dutilisation/

Cet article est diffusé et préservé par Érudit.

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche.

https://www.erudit.org/fr/

Narrative of a Practice for the Adaptation and Use of a Training tool for Evaluative Judgment Based on Concordance[1]

Récit de pratique sur l'adaptation et sur l'utilisation d'un dispositif de formation au jugement évaluatif basé sur la concordance

Relato prático sobre a adaptação e utilização de um dispositivo de formação do juízo avaliativo baseado na concordância

Marie-France Deschênes
ID ORCID : 0000-0002-7949-4466
Université de Montréal

Eric Dionne
ID ORCID : 0000-0003-3046-4386
Université d'Ottawa

Michelle Dorion
ID ORCID : 0000-0001-5173-5435
Université d'Ottawa

1. The French version was published in issue 46(2) 2023: https://doi.org/10.7202/1111098ar

Cette oeuvre est distribuée sous licence Creative Commons Attribution 4.0 International
http://creativecommons.org/licences/by/4.0/deed.fr

Key words: educational tool, evaluative judgment, learning assessment, learning by concordance, pedagogical innovation

Opportunities are rare for future teachers to test their interpretation of complex, uncertain, or ambiguous questions in learning assessment and to obtain subsequent formative feedback. This article proposes a pedagogical practice based on the adaptation and use of a Concordance of Judgment educational tool to exercise future teachers' evaluative judgment. The theoretical foundations of the educational tool, the principles of its design, as well as the contextual factors that have influenced its adaptation in an educational context are discussed. This analysis identifies benefits of using the tool to promote the development of teachers' evaluative judgment.

Mots clés : dispositif éducatif, évaluation des apprentissages, formation par concordance, innovation pédagogique, jugement évaluatif

Les futurs enseignants ont rarement l'occasion de tester leurs interprétations de questions complexes, incertaines ou ambiguës en évaluation des apprentissages et d'obtenir une rétroaction formative sur ces questions. Dans cet article, nous présentons un récit de pratique sur l'adaptation et sur l'utilisation d'un dispositif de formation au jugement évaluatif de futurs enseignants basé sur la concordance de jugement. Les assises théoriques du dispositif, les principes de sa conception ainsi que les facteurs contextuels ayant influencé son adaptation sont abordés. Un tel exercice nous permet de cerner quelques avantages de l'utilisation du dispositif pour favoriser le développement du jugement évaluatif en éducation.

Palavras chave: avaliação da aprendizagem, dispositivo educativo, formação por concordância, inovação pedagógica, juízo avaliativo

Os futuros professores raramente têm a oportunidade de testar as suas interpretações de questões complexas, incertas ou ambíguas em avaliação das aprendizagens e de receber feedback formativo sobre estas questões. Neste artigo, apresentamos uma narrativa de prática sobre a adaptação e o uso de um dispositivo de formação no juízo avaliativo de futuros professores baseado na concordância de juízo. Discutimos as bases teóricas do dispositivo, os princípios da sua conceção, bem como os fatores contextuais que influenciaram sua adaptação. Tal exercício permite identificar algumas vantagens do uso do dispositivo para promover o desenvolvimento do juízo avaliativo em educação.

Authors' note: Correspondence regarding to this article may be addressed to marie-france.deschenes@umontreal.ca

Introduction

All teachers are confronted with complex or ambiguous learning assessment situations that require them to weigh their pedagogical decisions and exercise evaluative judgment (Chaumont & Leroux, 2018; Dionne & Simões Forte, 2013; Jabůrek et al., 2022; Leroux & Bélair, 2015). Evaluative judgment, the professional judgment exercised for learning assessment (Maes et al., 2019; Maes et al., 2020; Mottier Lopez & Allal, 2008), is a fundamental competency teachers need to develop. Indeed, most learning assessment policies in Canada highlight the importance of evaluative judgment (Government of Ontario, 2010; Ministère de l'Éducation du Québec, 2003). However, opportunities for future teachers to test their evaluative judgements where pedagogical decisions include elements of complexity, ambiguity, or uncertainty are rare (Dionne & Simões Forte, 2013; Fives & Barnes, 2020; Smith, 2017).

In this article, we present a narrative of a pedagogical practice that adapts a training tool for evaluative judgment based on judgment concordance. This tool, translated from the French as the Concordance of Judgment Test in Learning Assessment, herein referred to as the CJTLA, was designed to develop teachers' competence for evaluating secondary pupils' learning during teacher training. The CJTLA includes 25 situations of evaluative practice in teaching (Dionne & Simões Forte, 2013). Learning evaluation is the only domain addressed in the test because it embodies the complexity and omnipresence of professional judgment in teaching (Laveault, 2005, 2008; Leroux & Bélair, 2015; Maes et al., 2019). In the following paragraphs, we discuss the concept of evaluative judgment in teaching and research findings that report teachers' continued discomfort with its application. Next, we describe the pedagogical approach associated with the use of the CJTLA educational tool. To structure this description, we draw on Cianciolo and Regehr's (2019) framework and propose a stratified analysis of an innovative pedagogical practice or intervention. This framework enables us to examine the interactions among the theoretical foundations underpinning the use of the CJTLA educational tool, the

principles of its design and adaptation, and the contextual factors that influence its use to meet the pedagogical intentions targeted. Through this exercise, we identified advantages of using the CJTLA to develop the evaluative judgment of future teachers. Finally, we discuss the limits of the use of the educational tool and ways to develop its use in the field of education.

Evaluative judgment in teaching: a competency to be developed by future teachers

Assessing pupil learning via a competency-based approach is a major challenge for teachers (Chaumont & Leroux, 2018; Leroux & Bélair, 2015; Tourmen, 2015). Learning assessments that adopt a competency-based approach call for authentic conditions in which pupils realize a complex task or a concrete production (Tardif, 2017). To assess learning based on this type of task, teachers must exercise their evaluative judgment with respect to pupils' efficient mobilization and use of resources and their level of the targeted competencies (Bélair, 2014; Scallon, 2015; Tardif, 2006, 2017).

According to Mottier Lopez and Allal (2008), evaluative judgment is a heuristic approach where the teacher gathers evidence provided by a pupil, which they then interpret to make an informed decision about the pupil's level of competency. To make these decisions, teachers rely on rigorous, transparent, fair, and equitable assessment procedures. These procedures must also be adapted to specific situations. Evaluative judgment can also be understood as an individual cognitive act by the teacher. This act is integrated into a communication network in a social and institutional context and respects the requirements of the school system, the culture of the establishment, the teacher's values, and their pedagogical strategies (Chaumont & Leroux, 2018; Mottier Lopez & Allal, 2008; Tourmen, 2014).

However, pedagogical decisions in learning evaluation are not exclusively the result of a rational process of data processing and analysis. A teacher's evaluative judgment is also influenced by their conceptions of teaching, their values, and their experiences (Laveault, 2008; Leroux & Bélair, 2015; Piot, 2008; Tourmen, 2015). Teachers rely on their intuition (Vanlommel et al., 2018) and individual representations of the very nature of competence (Tourmen, 2015). As a result, it is difficult to explain the motives, intentions, or circumstances supporting their decision. This is particularly true in ambiguous situations, when there is no univocal solution or consensus among the scientific community (Merle, 2012). The exercise of evaluative judgment is complex and inevitably subjective for teachers,

allowing the interference of cognitive biases in the evaluation process (Campbell, 2015; Merle, 2012, 2018). Incidentally, educational researchers have proposed evaluative practices to guide and minimize this subjectivity, such as the use of sufficient and relevant evidence of the development of competencies. Other researchers have proposed the use of criterion-referenced grids and evaluative practices that incorporate the principles of rigor, transparency and consistency, especially when assigning grades (Chaumont & Leroux, 2018; Leroux & Bélair, 2015; Seden & Svaricek, 2018).

Evaluative judgment is a dynamic and iterative process that presupposes an interpretation of evidence gathered and clues observed in a pupil's assignment that reflect targeted pedagogical intentions. Moreover, the practice also frequently involves the use of rigorous measurement tools to assess learning and accurately document that a level of competency is achieved (Leroux & Bélair, 2015; Loye & Fontaine, 2018). However, even the use of instrumented activity, or tools such as criterion-referenced grids, and competency frameworks, do not resolve the difficulties for teachers (Tourmen, 2014; Tourmen, 2015).

Complex, evaluative judgment frequently leads to teachers questioning themselves, and is hence a source of discomfort and uncertainty (Dionne & Simões Forte, 2013; Leroux & Bélair, 2015; Maes et al., 2019; Maes et al., 2020; Smith, 2017). Despite a growing body of scientific literature on this topic, studies have revealed that teachers feel persistent discomfort related to the exercise of evaluative judgment (Baribeau, 2015, 2020; Chaumont & Leroux, 2018; Leroux & Bélair, 2015; Maes et al., 2019; Maes et al., 2020). In a recent exploratory qualitative study, Maes et al. (2019) focused on the construction of evaluative judgment of supervisors during initial training of teachers during their internship. The authors conducted individual interviews with eight supervisors. The results of the study highlight the situated nature of evaluative judgment, shown by internship supervisors' consideration of singular aspects. The results also show that the final evaluative judgment is built on a series of partial and provisional judgments. Nevertheless, the majority of interviewed supervisors mentioned discomfort in exercising their evaluative judgment, particularly regarding the way the message is conveyed or when their judgment is questioned (Maes et al., 2019; Maes et al., 2020).

In her doctoral study, Baribeau (2015) conducted semi-structured interviews in the form of a narrative analysis of professional practice with teachers (n=12). The aim of this interpretative qualitative study was

to analyze self-reported assessment practices in summative decisions for certification of secondary pupils. The researcher's aim was to understand how these teachers constructed their evaluative practices. The results of the interviews were then presented to other teachers (n=10) to gain a more global perspective on the phenomenon concerned. The results of the study reveal obstacles and tensions in learning assessment practices adopted by teachers. The results show that these practices can be somewhat arbitrary with teachers relying on their appreciation of the pupil overall and their impressions of the pupil's engagement in learning. The researcher advances that this may be partially explained by the lack of professional training in learning assessment (Baribeau, 2015, 2020).

Educating future teachers in evaluative judgment is therefore essential in teacher education programs. Educational strategies such as theory courses and practicums are used to foster the development of future teachers' evaluative judgment (Leroux, 2019; Maes et al., 2019; Maes et al., 2020). However, despite their relevance, they seem insufficient for attenuating the persistent discomfort observed among teachers. In addition, during their training, future teachers are rarely faced with pedagogical decisions where there is uncertainty and ambiguity. More practice in these situations would foster the development of professional judgment (Deschênes et al., 2022; Dionne & Simões Forte, 2013; Smith, 2017). Mechanisms enabling future teachers to test their interpretations of complex pedagogical situations, reflect on issues raised in these situations, and obtain formative feedback are rare. In short, in the field of learning assessment, instruments that develop future teachers' competency for conducting evaluations remain insufficient (Baribeau, 2015, 2020).

An educational tool for exercising professional judgment in learning assessment

To foster the development of competency in learning assessment, we designed a training tool for evaluative judgment based on the judgment concordance, the CJTLA. We opted for the term 'educational tool' because this device aims to support learning through structured education characterized by the deliberate combination of online and classroom interventions (Boelens et al., 2017; Sacré et al., 2019). This conception is different from the use of pedagogical practice, which represents the various actions performed by the teacher, more or less consciously, to foster pupils' learning (Duget, 2014). The pedagogical intention behind this tool was the development of trainee teachers' competency in assessing secondary-pupil learning. In line

with Cianciolo and Regehr's (2019) framework for a stratified analysis of innovative pedagogical practice or intervention, we proceed by describing: (a) the theoretical foundations of the educational tool; (b) the structural aspects of the practice and the principles of the design; and (c) contextual factors influencing its adaptation in the field. This analysis helped us to determine whether the pedagogical practice meets the desired pedagogical intention and to draw plausible conclusions about the factors responsible for this result.

Theoretical foundations of the training tool

The CJTLA is an adaptation of the Script Concordance Test (SCT), a clinical reasoning assessment tool developed in the late 1990s for use in medical education, and more broadly in the health sciences (Dory et al., 2012; Lubarsky et al., 2011; Lubarsky et al., 2013). SCTs aim to measure the degree of concordance between respondents' answers to a test and experts' responses to the same questions. The SCT is based on script theory, which postulates that individuals cannot understand life situations without relying on their mental representations (Abelson, 1975; Schank & Abelson, 1977). Individuals react according to what seems most relevant to them, based on knowledge they have memorized in the form of models or schemas. A script is defined as mental representation that stereotypes pieces of situational information into units of meaning (Schank & Abelson, 1977). It is conceptualized as a finely organized knowledge network in the long-term memory (Abelson, 1975; Schank & Abelson, 1977).

In the 1980s, script theory was transposed to medical education to describe clinical reasoning, particularly with the illness script. Consistent with cognitivist foundations, the illness script involves associative links between the different types of knowledge, both theoretical and clinical, mobilized in the physician's clinical reasoning process. This mobilization of knowledge enables correct diagnosis, investigation, and treatment (Charlin et al., 2000; Schmidt et al., 1990). For example, the script for a myocardial infarction contains associative links between knowledge and the following data: (a) typical clinical signs (e.g., chest pain) and atypical signs; (b) precipitating and predisposing factors; (c) history of heart disease; (d) immediate evaluation and treatments needed; and (e) potential consequences. From a cognitive standpoint, illness scripts enable physicians to engage in hypothetico-deductive clinical reasoning. In concrete terms, physicians can compare the data from a clinical situation with their own scripts. This helps them to recognize patterns, similarities, or salient elements of the situation

to guide their clinical reasoning. They then search for additional data to minimize, reinforce, or prioritize clinical hypotheses. From this cognitivist point of view, expertise in clinical reasoning is linked not only to the depth of the physician's knowledge, but also to how knowledge is organized in their long-term memory (Charlin et al., 2000; Custers, 2015).

In the field of education, whether such scripts exist and how they are conceptualized in the teaching role is yet to be determined. We approach these questions using Mottier Lopez and Allal's (2008) conception of evaluative judgment, presented earlier. Their definition explains the presence of teachers' decision-making processes and the mobilization of knowledge derived from their expertise (experience and training). As in all reasoning processes, consciously or unconsciously, teachers generate hypotheses for pedagogical interventions by considering information they have acquired in their professional experience. They also use their expertise to identify decisive values and pedagogical principles to guide their judgment depending on the context. For example, a teacher's learning assessment script might include the following attributes: (a) principles of equity, fairness and transparency in learning assessment; (b) formalized institutional requirements such as rules and policies on learning assessment; (c) available resources; and (d) possible consequences of their decisions for the pupil such as motivation and academic progress (Chaumont & Leroux, 2018; Dionne & Simões Forte, 2013; Leroux & Bélair, 2015; Smith, 2017).

Constructs and traits are complex in both medicine and education, but while the former involves clinical reasoning, the latter involves evaluative judgment. The situations encountered in both professional contexts require individuals to interpret data, make decisions in uncertain situations, and deal with often missing, incomplete, or ambiguous information. Despite divergent characteristics, there are many similarities which suggest that the concordance tool, typical used in the medical education, can be adapted and used in an educational context (Deschênes et al., 2022; Dionne & Simões Forte, 2013; Smith, 2017).

Principles for designing the educational tool

To design a tool for the development of evaluative judgment through judgment concordance, we drew on our knowledge and followed approaches suggested in the scientific literature on using the SCT in the field of health sciences education (Dory et al., 2012; Lubarsky et al., 2013). In the following sections, we present the educational tool, its design principles, and its adaptation in an educational context.

The CJTLA and its design principles

The CJTLA, developed in 2008, is based on similarities observed between the complexity of professional judgment in teaching and the clinical reasoning process. The educational tool contains 25 vignettes to assess evaluative judgment, such as the quality of item writing for a measurement tool, planning an assessment, the concepts of validity and fidelity in an assessment, interpreting scores, and communicating results. A blueprint was also used to draw parallels between authentic teaching situations and components of the evaluative judgment process in the CJTLA items: planning, information gathering and interpretation, judgment, decision, and action.

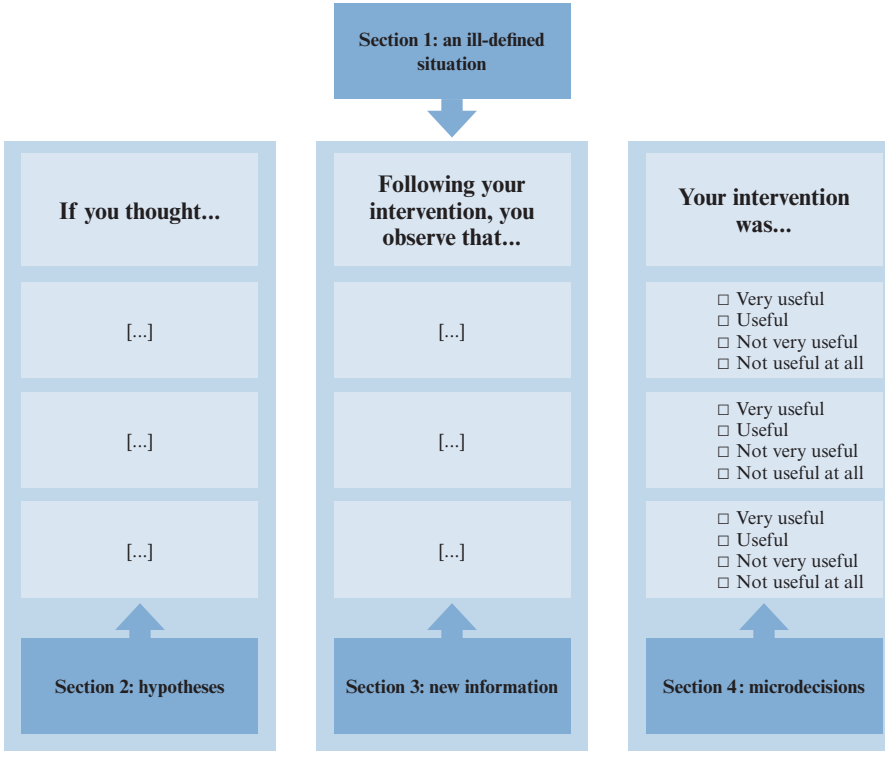
Each vignette in the CJTLA involves situations followed by items (n=35) to elicit micro-decisions[2] from student teachers. Each vignette is composed of four sections, as illustrated in Figure 1. Section 1 presents a short problem situation. The situation is designed to evoke a complex, ill-defined (Mayer & Wittrock, 2006) or ill-structured problem (Jonassen, 2011; Xun & Land, 2004). An ill-defined or ill-structured problem represents an authentic situation in practice where uncertainty or incompleteness persist in the decision-making process (Jonassen, 2011). There is no univocal situation or consensus for such situations in the scientific community for education, nor more generally in the social sciences (Voss, 1988). As a result, the situations in the educational vignettes are fragmentary, incomplete, or deliberately ambiguous. The choice of a common, representative, or high-stake problem is encouraged when drafting situations to maintain the significance or authenticity of the cognitive task assigned to the student teacher (Ashford-Rowe et al., 2014; Wiggins, 1993, 2011).

Sections 2 and 3 address items that lead student teachers to make micro-decisions that at least partially resolve the problematic situation. Hypotheses are presented (e.g., "If you are thinking about..."), followed by new information (e.g., "Following your intervention, you observe that..."). The hypotheses presented in Section 2 represent situations, plausible if taken independently, that student teachers might encounter in an educational environment. Of note, the new information revealed in Section 3 is designed to guide micro-decision making.

2. A micro-decision is a decision made with none of the information required to make a definitive judgment.

Section 4 allows student teachers to give their opinions on each of the hypotheses (e.g., "Your intervention was..."). When the educational tool is created, the wording of the answer choice categories can be adapted according to the targeted professional judgment. The relevance, acceptability, or usefulness of the intervention hypotheses are characteristics frequently used in the answer choices for the tool. The wording must be clearly drafted to prevent student teachers from making random choices. In addition, the possibility of providing a neutral answer (e.g., neither more nor less useful) is frequently avoided to avoid safe-harbour values on issues related to ethics, deontology, or professionalism. Experts with several years of experience in the teaching field answer all the items in the educational tool in advance. Experienced teachers are credible and likely to make appropriate judgments about the situations and hypotheses suggested in the vignettes. As partners in the design process of the educational tool,

Figure 1
Section of a vignette in the educational tool



they respond individually to the items without consulting colleagues or references (Dory et al., 2012; Lubarsky et al., 2013)[3]. This approach allows variability in the experts' choices. More than one option can therefore be considered in complex decisions when there is uncertainty and ambiguity.

Student teachers answering the questions in the educational tool can measure the degree of correspondence between their answers and experts' answers to the same questions. Originally, the student teachers' score for the SCT takes into account the answers previously given by experts. The more the student's response aligns with the experts' modal responses, the higher their score. Conversely, the further the selected response deviates from the experts' modal responses, the lower their score (Dory et al., 2012; Lubarsky et al., 2011). However, this method generates different scores depending on the number and characteristics of the experts consulted which poses a problem if the aim is to give student teachers an absolute score (Loye & Fontaine, 2018).

Adapting the educational tool

The design of the educational tool for exercising professional judgment in learning assessment reveals the transferability of the SCT from other professional fields. However, this raises questions about the construct (or trait) solicited for the situations and items in the educational tool.

In medical education, the illness script is a fairly algorithmic approach to decision-making based on key elements that categorize a person's manifestations or history regarding an illness. Take the example of a person suffering from a myocardial infarction. For problematic situations in education, teachers' knowledge networks do not seem to be based exclusively on knowledge useful for the categorization of educational situations. As mentioned by several authors (Chaumont & Leroux, 2018; Laveault, 2008; Maes et al., 2019; Maes et al., 2020; Mottier Lopez & Allal, 2008), teaching takes place in situated professional practice. Most pedagogical decisions are made with consideration for the school context, institutional rules, theory-based knowledge, and colleagues' opinions (Chaumont & Leroux, 2018; Laveault, 2008; Leroux & Bélair, 2015; Mottier Lopez & Allal, 2008).

Hence, there are two main challenges when drafting content for the educational tool: (a) the level of uncertainty required to meet the desired authenticity of the vignettes, and (b) the identification of the construct selected in the items. See Example 1 which illustrates our point.

3. Digital or E-learning environments enable experts to answer questions online and automatically generate feedback for student teachers.

Box 1. Example of a vignette in a learning assessment context

After assigning a summative assessment to your students, one of them asks you if they can submit an assignment they have already completed for another course that meets the requirements of your course.	
If you're thinking of...	**Following your intervention…**
1. ... allowing the pupil to submit this assignment.	... you correct the assignment and notice that the pupil has effectively mastered the content and is meeting expectations.
You consider that your intervention was:	☐ **Very useful** ☐ **Useful** ☐ **Not very useful** ☐ **Not useful at all**
2. ... allowing the pupil to hand in this assignment but asking them to add a short section explaining the differences between the way the subject was approached in the other course and in yours.	... the pupil gets a good grade on the assignment, but some pupils complain that this special permission is unfair.
You consider that your intervention was:	☐ **Very useful** ☐ **Useful** ☐ **Not very useful** ☐ **Not useful at all**
3. … not allowing the pupil to submit assignment already done.	... the pupil challenges your decision, explaining that the purpose of your assessment is to verify learning, not the context in which that learning took place.
You consider that your intervention was:	☐ **Very useful** ☐ **Useful** ☐ **Not very useful** ☐ **Not useful at all**

This problematic situation calls for micro-decisions in learning assessment using knowledge derived from the respondent experts' and student teachers' experience and training. In such authentic situations, teachers are called upon to demonstrate professionalism, to be familiar with the rules and principles of equity in learning assessment at their institution, and to work collegially with their peer group. However, pupils' questions and complaints combined with the context (e.g., institutional policies on learning assessment, professional group practices) influence their micro-decisions. In this and similar situations, the teacher's judgment is not exclusively egocentric but rather ecocentric, i.e. focused on the environment. It reflects a singular environment and accounts for the contextual, social material, institutional, and other aspects of the pedagogical situation. In short, "professional judgment is not isolated from the context in which it is exercised" (Leroux & Bélair, 2015, p. 99).

Contextual factors for using the training tool

The CJTLA has been used to educate students in a teaching program (Dionne & Simões Forte, 2013). In the following paragraphs, we discuss the contextual elements that influence the benefits observed in using the training tool.

Dionne and Simões Forte (2013) used the CJTLA with 72 student teachers in two groups enrolled in a Canadian university training program leading to secondary-level teaching. Each group was presented with four items during the session. During teaching practice, the vignettes of the tool were introduced in the classroom without feedback from experts. Using televoters[4], the student teachers indicated their responses for each of the individual items. They were asked to judge the usefulness of the proposed pedagogical interventions and the probability they would use them in an authentic context, i.e., in a real learning assessment situation. See Example 2 below.

Box 2. Vignette in learning assessment

For creating a complex task leading to learning certification, your colleague suggests you ask pupils to complete a self-assessment and peer assessment form within their work group.	
If you were thinking of...	**Following your intervention, you observe that ...**
...agreeing to ask the pupils to fill in the form and advising them that comments from their peers may influence their results.	...your pupils are confused about the weighting of this self-assessment by peers.
A. You consider that your intervention was	☐ **Very useful** ☐ **Useful** ☐ **Not very useful** ☐ **Not useful at all**
B. The probability that you performed this procedure is	☐ **Very likely** ☐ **Probable** ☐ **Unlikely** ☐ **Very unlikely**

4. These are devices that enable students to vote directly and synchronously during a course. They may be specialized devices or generic devices (e.g. tablets, laptops, phones) with internet connection.

A compilation of the student teachers' responses was then projected in graphs on a screen and a discussion between the class and the teacher was initiated. For each question, the student teachers were asked to state aloud any data they noticed was missing or ambiguous in the vignettes, as well as aspects they would change in the situation concerned. A qualitative-interpretive design was used, focusing on the analysis of their comments.

Following the use of the tool, Dionne and Simões Forte (2013) probed student teachers' level of appreciation via an anonymous questionnaire to judge the authenticity of the situations, the relevance of the items, and the benefits of the CJTLA for fostering the development of competence in learning assessment. A four-level Likert scale from "strongly disagree" to "strongly agree" was used. The student teachers were also given the opportunity to add explanatory comments to document their opinion of the CJTLA. The results showed that the student teachers found the items realistic (n = 62; 86%) for a teaching context. The results regarding the relevance of the items were mitigated: 48 student teachers (66%) found the items relevant, whereas 24 (33%) indicated that the items did not help them to understand the concepts associated with the assessment of learning. Finally, the majority of the student teachers reported that the use of the CJTLA motivated them to deepen their knowledge on the theme presented in the CJTLA (n = 57; 79%). In short, the results showed that student teachers favorably rated the use of the educational tool. They emphasized that the relevance of the content and the variety of situations in the vignettes prompted reflection on typical pedagogical decisions. They appreciated the cognitive effort required for resolving problematic situations during in-class discussion. The comments also testified to the student teachers' concern about understanding the situation presented more clearly and trying to define and find appropriate solutions. In this respect, the exchanges with the student teachers highlighted aspects of the evaluation contract proposed by Dionne and Chanelière (2022). The discussions helped them to identify deliberately missing, ambiguous, or incomplete information, which is a characteristic of the test or the concordance learning tool. In other words, student teachers were able to express their misunderstandings or identify contextual elements that were absent from the vignette that justified their response. The instrument captured a significant amount of information, but the qualitative phase proved crucial for exploring key elements of novice teachers' evaluative judgment.

In line with the socially situated practice of professional judgment in learning assessment, the results of this pedagogical practice suggest that the situations and items in the CJTLA vignettes should have slightly denser content than models currently used in medical education. The goal is to avoid polarizing experts' or student teachers' opinions in situations when they are insufficiently informed or documented to make a sound judgment of the relevance or acceptability of an educational intervention which may affect the process of responding to items. In other words, students' random responses illustrate a lack of understanding of the concepts or the desired direction of evaluative judgment in the situations presented to them, rather than uncertain decision making.

Conclusion

This article presents a pedagogical practice based on an educational tool that aims to foster the development of evaluative judgment in a teaching context. Cianciolo and Regehr's (2019) framework was used to examine the interactions among the theoretical foundations underpinning the use of the educational tool, the principles of its design, and the contextual factors influencing its adaptation to pedagogical intentions. The application of this framework supported our reflections on the benefits of the CJTLA for exercising professional judgment in learning assessment. Regarding the design of the tool, we observed the obvious presence of contextual elements that influenced evaluative judgment, affecting the writing and depth of the vignette content. The use of the tool enabled us to observe its relevance for future teachers' education on evaluative judgment, particularly in individual or group reflection (Dionne & Simões Forte, 2013).

Limitations and perspectives

This study is limited by the exploratory method underpinning the narrative of a practice. Future research would benefit from going beyond the contexts presented for using the educational tool and further exploring its use in research and teaching. In this regard, Smith (2017) conducted an exploratory study to understand the development of teachers and future teachers' judgment using the CJTLA. Semi-structured individual interviews were conducted via the think-aloud technique, a method of collecting verbal data from a person as they perform a complex cognitive process (Ericsson & Simon, 1980; Newell & Simon, 1972; Van Someren et

al., 1994). In Smith's (2017) study, participants (n = 12) described the key concepts elicited in CJTLA situations. They were also asked to judge the relevance of proposed hypotheses related to the pedagogical intervention and the probability that they would use this type of intervention in an educational context. The question pertaining to the probability they would use this type of intervention aimed to detect signs of social desirability due to their desire to show themselves in a favorable light (Crowne & Marlowe, 1960). The 12 participants were categorized according to three different profiles: (a) novice students, (b) experienced practicing teachers, and (c) teachers expert in learning assessment. The results of Smith's (2017) study showed that practitioners and experts were able to name key concepts in learning assessment better than novices. They were also more likely to identify missing data in the vignettes that would have been useful in refining their professional judgment. Smith (2017) recommended exploring the theoretical and ethical concepts elicited by the tool to model student teachers' cognitive processes. Using the tool during education could also help guide student teachers' evaluative practices and support the development of their professional judgment. Further experiments are needed to investigate the potential short- and long-term benefits of using an educational tool for the development of professional judgment in learning assessment.

Another limitation in the use of the tool is linked to the choice of experts and partners in its design. It is recommended to include credible people who are likely to make relevant judgments about the situations and hypotheses suggested in the vignettes of the educational tool (Dory et al., 2012; Lubarsky et al., 2011; Lubarsky et al., 2013). However, the criteria for determining the attributes or characteristics of the experts remain ambiguous and poorly documented. This begs the question regarding the criteria for specifying a teacher's expertise: experience, academic qualifications, or recognition of their credibility by their peers. It is difficult to identify the distinctive traits of education experts, because educational situations are characterized by the context.

Using instruments in teacher education to develop evaluative judgment is rare. Hence, the educational tool presented in this article is innovative, with many promising prospects for application. The first step is to create an online training course using CJTLA that incorporates the opinions of experts in the field of learning assessment. To date, CJTLA has been used to survey student teachers' level of appreciation and their cognitive processes. Asking experts to answer and comment on CJTLA items would

create formative feedback for student teachers to create an online educational tool. Like medical students, student teachers could benefit from automated feedback comprising both experts' answers and explanatory comments. Such comments would be a rich resource for student teachers' learning, particularly given the nuances and subtleties documented in the judgment process (Charlin et al., 2018; Charlin et al., 2021; Fernandez et al., 2023). Finally, other key resources on learning assessment could be consulted, such as formalized institutional requirements and key reference works. The aim of this approach is to guide student teachers' learning and synthesize the pedagogical intentions underlying their approach to problem-solving. Such an educational system would enable student teachers to reflect on various issues; obtain formative feedback on complex or uncertain questions they will face in professional practice; understand that, in practice, solutions can be ambiguous; and better apprehend the uncertainty of certain pedagogical decisions.

Despite these limitations, the stratified analysis of pedagogical practice performed in this article enabled us to identify the theoretical foundations and design principles of the tool and how it could be adapted to the educational context. This analysis also helped us refine the design of the educational tool to ensure it aligns with the principles of evaluative judgment. Identification of the difficulties encountered in writing the situations and items of the tool in an educational context reinforces the complex nature of professional judgment in teaching.

Proofreading: Caroline Lefour

Formatting: Emmanuel Gagnon

Portugue abstract: Eusébio André Machado

Original version (in French)

Received: March 23, 2022

Final version: January 17, 2024

Accepted: January 24, 2024

LIST OF REFERENCES

Abelson, R. P. (1975). Concepts for representing mundane reality in plans. In D. G. Bobrow and A. Collins (Eds.), *Representation and understanding: Studies in cognitive science* (pp. 273-309). Academic Press.

Ashford-Rowe, K., Herrington, J. & Brown, C. (2014). Establishing the critical elements that determine authentic assessment. *Assessment & Evaluation in Higher Education*, 39(2), 205-222. https://doi.org/10.1080/02602938.2013.819566

Baribeau, A. (2015). *Analyse des pratiques d'évaluation d'enseignants du secondaire IV et V dans des décisions sommatives de certification des apprentissages des élèves* [Thèse de doctorat, Université du Québec à Trois-Rivières et Université du Québec à Montréal].

Baribeau, A. (2020). La professionnalité de l'agir évaluatif de l'enseignant du secondaire dans le contexte québécois. *Administration et Éducation*, 165, 233-239. https://doi.org/10.3917/admed.165.0233

Bélair, L. (2014). Évaluer ce qu'ils ont appris. Dans L. Ménard et L. St-Pierre (dir.), *Se former à la pédagogie de l'enseignement supérieur* (p. 355-380). Association québécoise de pédagogie collégiale (AQPC).

Boelens, R., De Wever, B. & Voet, M. (2017). Four key challenges to the design of blended learning: A systematic literature review. *Educational Research Review*, *22*, 1-18. https://doi.org/10.1016/j.edurev.2017.06.001

Campbell, T. (2015). Stereotyped at Seven? Biases in Teacher Judgement of Pupils' Ability and Attainment. *Journal of Social Policy*, *44*(3), 517-547. https://doi.org/10.1017/S0047279415000227

Charlin, B., Deschênes, M.-F., Dumas, J.-P., Lecours, J., Vincent, A.-M., Kassis, J., Guertin, L., Gagnon, R., Robert, D., Foucault, A., Lubarsky, S. & Fernandez, N. (2018). Concevoir une formation par concordance pour développer le raisonnement professionnel : quelles étapes faut-il parcourir ? *Pédagogie Médicale*, *19*, 143-149. https://doi.org/10.1080/0142159X.2021.1900554

Charlin, B., Deschênes, M.-F. & Fernandez, N. (2021). Learning by concordance (LbC) to develop professional reasoning skills: AMEE Guide No. 141. *Medical Teacher*, 614-621. https://doi.org/10.1080/0142159X.2021.1900554

Charlin, B., Tardif, J. & Boshuizen, H. P. A. (2000). Scripts and medical diagnostic knowledge: Theory and applications for clinical reasoning instruction and research. *Academic Medicine*, *75*(2), 182-190. https://doi.org/10.1097/00001888-200002000-00020

Chaumont, M. & Leroux, J. L. (2018). Le jugement évaluatif: subjectivité, biais cognitifs et postures du professeur. *Pédagogie collégiale*, *31*(3), 27-33.

Cianciolo, A. T. & Regehr, G. (2019). Learning Theory and Educational Intervention: Producing Meaningful Evidence of Impact Through Layered Analysis. *Academic Medicine*, *94*(6). https://journals.lww.com/academicmedicine/Fulltext/2019/06000/Learning_Theory_and_Educational_Intervention_.31.aspx

Crowne, D. P. & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of consulting psychology*, *24*(4), 349-354.

Custers, E. J. (2015). Thirty years of illness scripts: Theoretical origins and practical applications. *Medical Teacher*, *37*(5), 457-462. https://doi.org/10.3109/0142159X.2014.956052

Deschênes, M.-F., Tremblay, K., Pelletier, I., Charlin, B. & Fernandez, N. (2022). Expérience d'enseignants à la réalisation d'un dispositif numérique de formation basée sur la concordance de jugement. *Formation et Profession*, *30*(2), 1-13. https://doi.org/10.18162/fp.2022.704

Dionne, E. & Chanelière, M. (2022). Du contrat didactique au contrat évaluatif : une grille de lecture pour mieux appréhender l'évaluation des apprentissages en pédagogie médicale. *Pédagogie médicale*, *23*(1), 85-92. https://doi.org/10.1051/pmed/2021030

Dionne, É. and Simões Forte, L. A. (2013). Le test de jugement professionnel en évaluation des apprentissages (CJTLA) : un outil visant à mesurer des compétences avec des télévoteurs. In *Actes du Colloque scientifique international sur les TIC en éducation : bilan, enjeux actuels et perspectives futures*, Montréal.

Dory, V., Gagnon, R., Vanpee, D. & Charlin, B. (2012). How to construct and implement script concordance tests: insights from a systematic review. *Medical Education*, *46*(6), 552-563. https://doi.org/10.1111/j.1365-2923.2011.04211.x

Duget, A. (2014). *Les pratiques pédagogiques en première année universitaire : description et analyse de leurs implications sur la scolarité des étudiants* [Doctoral dissertation, University of Burgundy].

Ericsson, K. A. & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*, 215-251. https://doi.org/10.1037/0033-295x.87.3.215

Fernandez, N., Deschênes, M.-F., Akremi, H., Lecours, L., Jobin, V. & Charlin, B. (2023). What can Designing Learning-by-Concordance Clinical Reasoning Cases Teach Us about Instruction in the Health Sciences? *Perspectives on Medical Education*, *12*(1), 160-168. https://doi.org/10.5334/pme.898

Fives, H. & Barnes, N. (2020). Navigating the complex cognitive task of classroom assessment. *Teaching and Teacher Education*, *92*, 103063. https://doi.org/10.1016/j.tate.2020.103063

Gouvernement de l'Ontario. (2010). *Faire croitre le succès : évaluation et comunication du rendement des élèves fréquentant les écoles de l'Ontario*. Ministère de l'Éducation de l'Ontario (MEO). http://www.edu.gov.on.ca/fre/policyfunding/growSuccessfr.pdf

Jabůrek, M., Cígler, H., Valešová, T. & Portešová, Š. (2022). What is the basis of teacher judgment of student cognitive abilities and academic achievement and what affects its accuracy? *Contemporary Educational Psychology*, *69*, 102068. https://doi.org/10.1016/j.cedpsych.2022.102068

Jonassen, D. H. (2011). *Learning to solve problems: A handbook for designing problem-solving learning environments*. Routledge.

Laveault, D. (2005). Le jugement professionnel de l'enseignant : quel impact sur l'acte d'évaluer ? *Mesure et évaluation en éducation*, *28*(2), 93-114. https://doi.org/10.7202/1087226ar

Laveault, D. (2008). Le jugement professionnel : foyer de tensions et de synergies nouvelles en évaluation scolaire. *Revue suisse des sciences de l'éducation*, *30*(3), 483-500. https://doi.org/10.25656/01:4230

Leroux, J. (2019). Les multiples facettes du travail des formateurs de stagiaires : quels enjeux et quels défis? *Phronesis*, *8*(1-2), 111-126. https://doi.org/10.7202/1066588ar

Leroux, J. & Bélair, L. (2015). Exercer son jugement professionnel en enseignement supérieur. Dans J. Leroux (dir.), *Évaluer les compétences au collégial et à l'université : un guide pratique* (p. 67-104). Association québécoise de pédagogie collégiale (AQPC).

Loye, N. & Fontaine, S. (2018). S'instrumenter pour évaluer. *Pédagogie médicale*, *19*(2), 95-107. https://doi.org/10.1051/pmed/2019018

Lubarsky, S., Charlin, B., Cook, D., Chalk, C. & Van der Vleuten, C. (2011). Script concordance testing: a review of published validity evidence. *Medical Education*, *45*(4), 328-338. https://doi.org/10.1111/j.1365-2923.2010.03863.x

Lubarsky, S., Dory, V., Duggan, P., Gagnon, R. & Charlin, B. (2013). Script concordance testing: from theory to practice: AMEE guide no. 75. Medical Teacher, 35(3), 184-193. https://doi.org/10.3109/0142159X.2013.760036

Maes, O., Colognesi, S. & Van Nieuwenhoven, C. (2019). Le processus de construction du jugement évaluatif par les superviseurs de stage en enseignement. *Mesure et évaluation en éducation*, *42*(1), 35-61. https://doi.org/10.7202/1066597ar

Maes, O., Van Nieuwenhoven, C. & Colognesi, S. (2020). La dynamique de construction du jugement évaluatif du superviseur lors de l'évaluation de stages en enseignement. *Canadian Journal of Education*, *43*(2), 522-547. https://www.jstor.org/stable/26954697

Mayer, R. E. & Wittrock, M. C. (2006). Problem solving. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (2e ed., pp. 287-303). Lawrence Erlbaum.

Merle, P. (2012). L'évaluation des élèves. Une modélisation interactionniste des pratiques professorales. Dans L. M. Lopez & G. Figari (dir.), *Modélisations de l'évaluation en éducation : questionnements épistémologiques* (p. 81-96). De Boeck Supérieur.

Merle, P. (2018). Normes et biais généraux d'évaluation. Dans P. Merle (dir.), *Les pratiques d'évaluation scolaire : historique, difficultés, perspective*. Presses Universitaires de France.

Ministère de l'Éducation du Québec. (2003). *Politique d'évaluation des apprentissages : Formation générale des jeunes, formation générale des adultes, formation professionnelle*. Ministère de l'Éducation du Québec. https://www.education.gouv.qc.ca/fileadmin/site_web/documents/dpse/evaluation/13-4602.pdf

Mottier Lopez, L. & Allal, L. (2008). Le jugement professionnel en évaluation : un acte cognitif et une pratique sociale située. *Revue suisse des sciences de l'éducation*, *30*(3), 465-482. https://doi.org/10.24452/sjer.30.3.4798

Newell, A. & Simon, H. A. (1972). Human problem solving. Prentice-Hall

Piot, T. (2008). La construction des compétences pour enseigner. *McGill Journal of Education/Revue des sciences de l'éducation de McGill*, *43*(2), 95-110. https://doi.org/10.7202/019577ar

Sacré, M., Lafontaine, D. & Toczek, M.-C. (2019). Liens entre les composantes des dispositifs d'enseignement hybride et les performances des étudiants de l'enseignement supérieur : une revue systématique. *Mesure et évaluation en éducation*, *42*(3), 109-152. https://doi.org/10.7202/1074105ar

Scallon, G. (2015). *Des savoirs aux compétences : exploration en évaluation des apprentissages*. De Boeck Supérieur.

Schank, R. & Abelson, R. (1977). *Script, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*. Lawrence Erlbaum Associates.

Schmidt, H. G., Norman, G. R. & Boshuizen, H. P. (1990). A cognitive perspective on medical expertise: theory and implication. *Academic Medicine*, *65*(10), 611-621.

Seden, K. & Svaricek, R. (2018). Teacher subjectivity regarding assessment: Exploring English as a foreign language teachers' conceptions of assessment theories that influence student learning. *CEPS Journal*, *8*(3), 119-139. https://doi.org/10.25656/01:16008

Smith, J. (2017). *Application d'un test de concordance : étude exploratoire du développement de la compétence à évaluer les élèves chez les enseignants* [Mémoire de maîtrise, Université d'Ottawa]. http://hdl.handle.net/10393/36510

Tardif, J. (2006). *L'évaluation des compétences: documenter le parcours de développement*. Chenelière éducation.

Tardif, J. (2017). Des repères conceptuels à propos de la notion de compétence, de son développement et de son évaluation. Dans P. J. Marianne, Tardif & G. François (dir.), *Organiser la formation à partir des compétences. Un pari gagnant pour l'apprentissage dans le supérieur* (p. 15-37). De Boeck Supérieur.

Tourmen, C. (2014). Contributions des sciences de l'éducation à la compréhension de la pratique évaluative. *Politiques et management public*, *31*(1), 69-85. http://journals.openedition.org/pmp/6970

Tourmen, C. (2015). L'évaluation des compétences professionnelles : apports croisés de la littérature en évaluation, en éducation et en psychologie du travail. *Mesure et évaluation en éducation*, *38*(2), 111-144. https://doi.org/10.7202/1036765ar

Van Someren, M., Barnard, Y. & Sandberg, J. (1994). Studying the content of cognitive processes. In M. Van Someren, Y. Barnard & J. Sandberg (Eds.), *The Think Aloud Method: A practical guide to modelling cognitive processes* (pp. 13-26). Academic Press.

Vanlommel, K., Van Gasse, R., Vanhoof, J. & Petegem, P. V. (2018). Teachers' high-stakes decision making. How teaching approaches affect rational and intuitive data collection. *Teaching and Teacher Education*, *71*, 108-119. https://doi.org/10.1016/j.tate.2017.12.011

Voss, J. F. (1988). Problem solving and reasoning in ill-structured domains. In *Analysing everyday explanation: A casebook of methods* (pp. 74-93). SAGE Publications.

Wiggins, G. (1993). Assessment: Authenticity, Context, and Validity. *The Phi Delta Kappan*, *75*(3), 200-214. http://www.jstor.org/stable/20405066

Wiggins, G. (2011). A True Test: Toward More Authentic and Equitable Assessment. *The Phi Delta Kappan*, *92*(7), 81-93. https://doi.org/10.1177/003172171109200721

Xun, G. E. & Land, S. M. (2004). A conceptual framework for scaffolding ill-structured problem-solving processes using question prompts and peer interactions. *Educational Technology Research and Development*, *52*(2), 5-22. https://doi.org/10.1007/BF02504836