Research Article https://doi.org/10.12973/ijem.8.3.479

International Journal of Educational Methodology
Volume 8, Issue 3, 479 - 491.
ISSN: 2469-9632
https://www.ijem.com/

Observational Record and Self-Report of Teacher-Student Performance in High School Lessons

Aldo Bazán-Ramírez*
Universidad Nacional Federico Villarreal, PERU

Néstor Miguel Velarde-Corrales
Universidad Pedagógica Nacional, MEXICO

María Elena Rodríguez-Pérez
Universidad de Guadalajara, MEXICO

Jorge Guerrero-Barrios
Universidad Nacional Autónoma de México, MEXICO

Roberta Brita Anaya-González
Universidad Nacional de San Cristóbal de Huamanga, PERU

Received: January 19, 2022 ▪ Revised: March 2, 2022 ▪ Accepted: June 20, 2022

Abstract: There are different strategies to analyze teacher and student performance when they interact with each other in class. The most used strategies are direct observation and verbal reports. Even though what is observed or reported depends on theoretical frameworks regarding didactic interactions, these must be related to teacher functions such as supervision, providing feedback, and evaluation of student performance. In this study, instruments for observational recording and verbal reports were developed and validated considering teacher functions and their student performance counterparts to compare the degree of correspondence or divergence between data gathered from both strategies. 135 students enrolled in a science class and their teachers participated. The class was taught in a public high school located in center/south Mexico. Classes were videotaped and the corresponding observational records were analyzed. Two months later, verbal reports were administered to students and teachers. Coincidences and differences that were found revealed that observer-observe interaction is conditioned by social norms. These results were interpreted considering the notion of silence as a communicative element.

Keywords: *Didactic interaction, direct observation methodology, verbal self-reports, teacher performance criteria, student performance criteria.*

To cite this article: Bazán-Ramírez, A., Velarde-Corrales, N. M., Rodríguez-Pérez, M. E., Guerrero-Barrios, J., & Anaya-González, R. B. (2022). Observational record and self-report of teacher-student performance in high school lessons. *International Journal of Educational Methodology*, *8*(3), 479-491. https://doi.org/10.12973/ijem.8.3.479

Introduction

In a school context, during the development of a class, the interaction between the students and the object of study is mediated by the teacher in accordance with the level of aptitude and competence development the students should attain based on certain achievements and specified criteria. This didactic practice reflects the circumstances and the characteristics of behavior that are desired from the apprentice, the achievement itself, and the fulfillment of achievement criteria (Ribes, 2008). In order to analyze the teacher-student-object of study relationship, teacher performance and student performance can be characterized separately when functionally interacting in class. This way, the behavior of the agents and the educational processes in a teaching/learning situation can be experimentally described and analyzed (Ibáñez, 2007).

Scholars dedicated to educational processes and, particularly, didactic interactions (relations between professor-student-object of study) have characterized actor performance during a didactic interaction in different ways. Likewise, based on the characterization, a diverse number of strategies have been derived for its study; the most frequent being the analysis of observational records and the use of self-report questionnaires. A tendency in the characterization and measurement of teacher performance and interactive processes in the classroom comes from educational approaches for measurement and evaluation. In general, these approaches are oriented by a substantive theory and are centered around the type of performance to be identified and validated through statistical procedures with advanced software (Martínez & Moreno, 2002; Mislevy, 1993).

* **Corresponding author:**
Aldo Bazán-Ramírez, Universidad Nacional Federico Villarreal, Peru. ✉ abazanramirez@gmail.com

Behavioral approaches have also characterized teacher and/or student performance in didactic interactions (Carpio et al., 1998; Hugh-Pennie et al., 2018; Irigoyen et al., 2011; Silva et al., 2014; Vargas, 2020). Some studies have drifted towards the use of observational methodologies to register and analyze didactic performance in different levels of education based on behavioral categories (Bazán-Ramírez et al., 2009; Borges & Falcón, 2018; Díaz et al., 2015; Mares et al., 2020; Velarde-Corrales & Bazán, 2019).

In general, regardless of the conceptual frames, the categories for the didactic performance of a teacher converge in six groups: a) The explanation the teacher gives the students regarding the learning objectives and the criteria for performance or activity. b) The degree or level of functional structurization of the student-object of study relation mediated by teacher performance. This process has also been characterized by cognitive perspectives such as cognitive activation (Gitomer, 2019; Simpson & Bester, 2017; Van der Lans et al., 2018). c) The way in which the professor-student-object of study relation is structured which can be in the form of explanation, illustration, or demonstration by the teacher. d) Student supervision or accompaniment during the learning process. e) Feedback based on the expected achievements and the criteria for their execution, and f) Progressive and final evaluation of student skills and competencies.

These categories for teacher performance in a didactic interaction suppose the need of including categories for student performance as well. Within a behavioral perspective, teacher and student performance categories are not usually analyzed simultaneously in secondary and higher levels of education; except for the work conducted by Mares et al. (2020), as well as Velarde-Corrales and Bazán (2019) along with the models that Ibáñez (1999) and Morales et al. (2017) proposed. Strategies for studying the teacher-student relation in a didactic interaction have varied since the use of checklists, going from interviews and the analysis of observational records to the application of self-report questionnaires. For this project, observational records and questionnaires were used.

Studies with an observational methodology focus on comprehending and explaining the practice of teaching, instructional processes, and the quality of education by using a system that registers teacher and student behavior during class. The observational record has been used in different investigations based on both cognitive perspectives and behavioral approaches (Bell et al., 2019; Borges et al., 2016; Gitomer, 2019; Joyce et al., 2018; Peralta & Roselli, 2015; Ruiz et al., 2020; Velarde-Corrales & Bazán, 2019). Thus, studies have reported using observational records of teacher performance categories (Borges & Falcón, 2018; Borges et al., 2016; Díaz et al., 2015) and categories or criteria of teacher and student performance in interactive teaching/learning processes (Galindo et al., 2017; Velarde-Corrales & Bazán, 2019).

In the case of verbal reports, there is a vast tradition in the study of teacher behavior during the development of a lesson in secondary and higher levels of education. Teacher performance in didactic interactions is commonly evaluated by the students through assessment surveys or questionnaires (Bazán-Ramírez et al., 2021; Chan, 2018; Liu & Cohen, 2021; Scherer et al., 2016; Üstünlüoglu & Güngör-Culha, 2012). To a lesser extent, evaluations with self-report questionnaires have been developed to concurrently inquire about teacher and student performance in didactic interactions (Bazán-Ramírez & Velarde-Corrales, 2021; Bazán-Ramírez et al., 2022).

These instruments of self-report attempt to identify the tendency or frequency of occurrence of certain behaviors both from the teacher and the student during the didactic interactions which happen shortly before the moment of their assessment. Studies reported under this perspective have mostly centered around the measurement of diverse categories of teacher performance via questionnaires and the validation of its content and hypothetical constructs (Chan, 2018; Grammatikopoulos et al., 2015; Krijgsman et al., 2019; Nasser-Abu, 2017).

The correspondence between data about the occurrence of an expressed behavior in a natural situation which is obtained by behavioral records and data about tendencies of behavior evaluated with self-reports is an issue that requires attention from researchers in education given that both procedures allow the gathering of valuable information which can be complementary. While a self-report questionnaire provides indirect access to information about the tendencies of teacher performance (what students and professors describe), an observation or an observational record allows the identification of an occurrence and the sequence of performances that are being displayed (which are indicators of theoretically established categories) in a natural situation, before extraneous variable control to the implemented experimental design.

Various authors have reported the importance, and limitations, of utilizing observational instruments and self-reports in educational practice research (Duckworth & Yeager, 2015; Salinas, 2017). Using these instruments in a complementary way is suggested. For example, in the assessment of teaching practices within inclusive education in Mexico (García et al., 2015), and in college-level virtual teaching in Argentina (Salinas, 2017), correspondence between the evaluated categories in both instruments has been reported.

On the contrary, a study at a primary level of education with North American professors, who took a course in professional development, reported no significant correlation between the teacher's self-report about their knowledge in mathematics for teaching purposes and the direct observation evaluations of their performance (Copur & Thacker, 2021). Apparently, there is less correspondence between the information obtained using questionnaires and

observational instruments when self-reporting is about beliefs and opinions regarding certain aspects of behavior. In contrast, when behavioral competencies are evaluated, the information that is collected using observational records and questionnaires presents greater correspondence (Corral-Verdugo, 1997).

Based on the literature that has been presented, the aim of this study was to compare the information about teacher-student performance criteria that was obtained using an observational record system and a self-report questionnaire which was administered to students and professors in five different high school science classes. For the aforementioned, an observational record system was validated using five categories for the teacher performance criteria and five more categories for student performance. Subsequently, a self-report questionnaire was constructed transforming the behavioral criteria of the observational record into a Likert-type questionnaire. Teacher performance and student performance were both considered resulting in the construction of two questionnaires. Both self-report questionnaires were validated. Finally, the data that was obtained from the behavioral records and both questionnaires were analyzed and compared.

Methodology

Research Design

A non-experimental design with a mixed methodology was implemented. To gather information two sources were considered: observational records from different classes (Anguera et al., 2011) and self-report questionnaires.

Sample

In this study, 135 students (52 men and 83 women) and five teachers (four men and a woman) from the area of sciences participated. They were in their second semester of high school in a public college in Mexico, specifically the August – December semester of 2017. The courses where the evaluations took place were Mathematics (35 students and the professor), Biology (37 students and the professor), Physics (32 students and the professor), Instrumentation (31 students and the professor), and Biochemistry (31 students and the professor).

The average age of the students was 17 years old (SD = 1 year) and the average age of the professors was 41 years (SD = 12.5 years). All teachers had a postgraduate degree and more than 3 years of experience as a teacher at secondary levels of education.

Categories for the Teacher-Student Performance Criteria

Five pairs of didactic performance criteria were used. These were validated and reported by Velarde-Corrales and Bazán (2019) for the analysis of didactic interactions in science in a secondary level of education school. The criteria are as follows: Exploration of competencies, Explicitation of criteria, Illustration, Feedback, and Evaluation. This five-criteria proposal, the teacher performance criteria, and the corresponding student performance criteria were based on the five-criteria didactic performance model presented by Carpio et al. (1998), as well as the seven-criteria model presented by Irigoyen et al. (2011) and Silva et al. (2014).

Instruments

Observational Record.

Five records of five classes were taken into account based on the Observational System for the Analysis of Didactic Interactions or “Sistema Observacional Para Analizar Interacciones Didácticas (SOPAID)” in Spanish (Velarde-Corrales & Bazán, 2019). The reliability or agreement index among observers was good (Kappa coefficient = 0.82). The observational instrument included five areas of teacher performance (Exploration of competencies, Explicitation of criteria, Illustration, Feedback, and Evaluation) and five areas of student performance (Student adjustment to the exploration of competencies, Identification of criteria, Adjustment to the linguistic model and illustration, Participation – Student Adjustment to feedback, Evaluation and Application). Table 1 exposes an example of the SOPAID record categories used in this study.

Table 1. Example of Indicators and Dimensions of Teacher and Teacher Performance

Teacher performance		**Student performance**	
Explicitation of criteria	Indicators	Identification of criteria	Indicators
Explains the parameters that the student must satisfy according to the didactic criteria.	A. Describes the necessary requirements to carry out classwork. B. Describes the actions the student must achieve to fulfill certain activity. C. Explains what the student must do to complete an exercise in class.	Reproduces and/or recreates the criteria of the course or the lesson; asks "what" questions and how they must adjust to the criteria.	A. Complies with the criteria that the professor indicates for classwork. B. Carries out the activity complying with the criteria the teacher explained. C. Complies with the criteria the professor mentioned while working on an exercise in class.

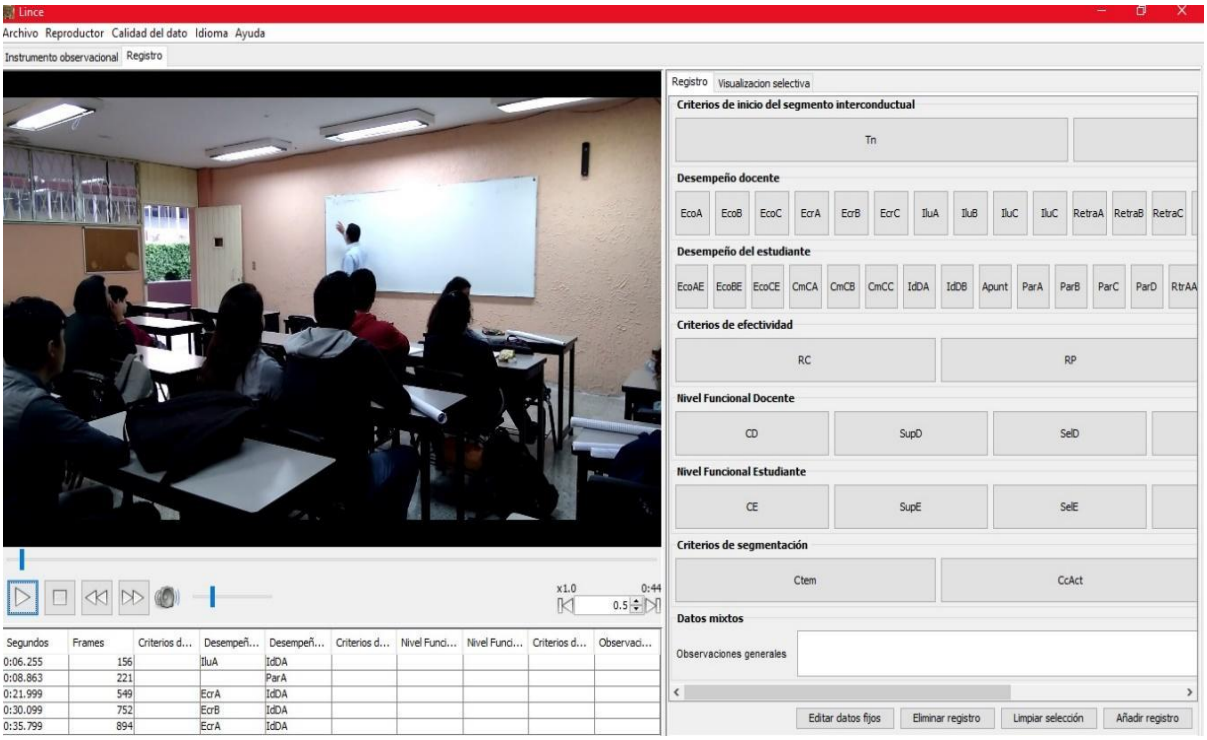

Figure 1. Example of the Record of Interactions in LINCE Software

Figure 1 shows an example of the programming for the didactic interactions record using the system for observation and records SOPAID through LINCE software. The example exhibits the Explicitation of Criteria – Identification of Criteria performance.

Didactic Interactions Questionnaire (CID, in Spanish).

In order to identify the student's ratings of teacher performance and of their own performance in a science lesson, two questionnaires were designed; both were based on the five dimensions of the SOPAID, along with their indicators from the record system. One questionnaire explored the student's assessment of teacher performance while the other consisted of the student's self-assessment of their own performance during the didactic interactions of the corresponding course. The questionnaires demanded Likert-type responses with these four options: 0 = Never, 1 = Seldom, 2 = Regularly, and 3 = Always.

Questionnaire for Student Assessment of Teacher Performance.

In Table 2, the 18 items that make up the questionnaire for teacher performance are presented and organized according to the theoretically corresponding performance criteria.

Table 2. Sentences for Each Criterion of Teacher Performance

Code	Items (statements)
Exploration of competencies	
TExplor1	The professor explores my knowledge of the subject at the beginning of the class.
TExplor2	The professor presents problems for me to solve before starting a topic.
TExplor3	The professor asks about concepts related to the topic before explaining it.
Explicitation of criteria	
TCriter4	The professor describes the requirements that will be needed to perform certain classwork.
TCriter5	The professor explains the criteria that will be required to perform an activity.
TCriter6	The professor explains what I need to do to complement an exercise done in class.
Illustration	
TIllust7	When explaining in class, the criteria that we must achieve are clear.
TIllust8	The professor clearly explains the topic of the lesson.
TIllust9	The professor exemplifies the assignment.
TIllust10	The professor solves problems in front of the students based on the topic at hand.
Feedback	
TFeedb1	The professor corrects me when I do not properly execute an activity in class.
TFeedb2	The professor makes me see what I did wrong and teaches me different ways in which I can solve the problem.
TFeedb13	The professor teaches me different ways in which I can satisfy the criteria established for the class activities.
TFeedb4	The professor checks and grades the assigned exercises and homework.
TFeedb5	The professor checks, corrects, and gives me notes on how to improve my work on assigned exercises and homework.
Evaluation	
TEvalu16	The professor assesses my theoretical and conceptual knowledge as well as the fundamentals of the course.
TEvalu17	The professor assesses the applied nature of the subject matter and presents problem-solving activities that can be derived from the course.
TEvalu18	The professor assesses my capacity to integrate the knowledge of different courses and this course.

This questionnaire went through confirmatory factor analysis (CFA), using the EQS 6.4 program, with a robust estimation method of maximum-likelihood (ML). The CFA for the student's assessment of teacher performance presents an acceptable statistical goodness-of-fit ($p < 0.001$; CFI = 0.91 and RMSEA = 0.07) and it confirmed the five dimensions specified in the instrument: Exploration of competencies, Explicitation of criteria, Illustration, Feedback, and Evaluation, as well as their respective indicators. The statistics program suggested that item 15, The professor checks, corrects, and gives me notes on how to improve my work on assigned exercises and homework, also loads in factor 5 (Evaluation). However, factorial loading is greater in factor 4 (Feedback), hence item 15 can be considered as an indicator of Feedback. On the other hand, the covariance between the five factors varied between 0.36 and 0.88 but, in general, there is divergence among factors (constructs).

Self-Assessment Questionnaire for Student Performance.

A self-assessment questionnaire for student performance was elaborated considering five criteria. Criteria considered the following aspects: Previous competencies (Adjustment to the exploration of competencies), Compliance with criteria (Identification of criteria), Illustration (Adjustment to the linguistic mode and illustration), Participation and Student Adjustment to feedback, and Evaluation and application.

This questionnaire also went through confirmatory factor analysis (CFA), using the EQS 6.4 program, with a robust estimation method of maximum-likelihood (ML). The resulting CFA model confirmed the five dimensions of student performance criteria with excellent statistical goodness-of-fit ($p = 0.16$; CFI = 0.97 and RMSEA = 0.03), obtaining acceptable convergent validity of construct and excellent divergent (discriminant) validity of construct. Table 3 presents the 17 items that resulted from the CFA organized in five dimensions or criteria for student performance for their self-assessment. This table already excludes an item that resulted in low statistical significance.

Table 3. Sentences for Each Criterion of Performance for Student Self-Assessment

Code	Items (statements)
Explored and existing competencies	
SExplor1	I respond or comment when the professor asks me or asks the class questions before starting a new topic.
SExplor2	I adequately solve the problems the professor presents before starting a new topic.
SExplor3	I respond to the questions the professor poses about new topics.
I comply with the established criteria	
SCriter4	I comply with the criteria the professor presents for carrying out classwork.
SCriter5	I perform an activity in compliance with the criteria the professor already explained.
SCriter6	I comply with the criteria the professor mentions when working on a class exercise.
Illustration	
SIllust9	I work on class exercises considering the model the teacher presented.
SIllust10	I solve the problems that the professor presents in class.
Participation – Feedback	
SParti11	I intervene in class to complete the topic that is being presented.
SParti12	I share my opinion about the topic of the lesson.
SParti13	I contribute to my classmate's comments.
SParti14	I ask questions about the topic that is being developed in class.
Evaluation (achievement)	
SEvalu7	I change my actions to comply with the criteria of the lesson or assignment.
SEvalu8	I comply with the criteria established for class activities.
SEvalu15	I comply with class assignments.
SEvalu16	I carry out the exercises that are related to the topic that is being developed.
SEvalu17	In my actions related to the lesson, I can identify if I have done my job well.

Equipment and Materials

LINCE software (version 1.4) was used to analyze the video material. A behavioral catalog specified the conditions in which the interactive segment began. It also described the categories and indicators used to identify interactions and evaluate teacher and student performances.

Ethical Considerations

Ethical considerations were covered with an informed consent signature for participants of legal age. In the case of participants that were minors, their parents received a consent and permission form which specified aspects of the student's participation in the investigation and had to be signed in agreement by both parents and students. The same protocol was implemented for the teachers that participated in the investigation with an informed consent form which they signed after reading and asking questions; they were informed and accepted their participation in the investigation.

Data Collection

At first, the data from the observational records registered in the LINCE program was gathered. The records came from the lessons that were videotaped when the students were studying during the second month of their second semester of school. With these records at hand, the frequencies of occurrence of action according to each criterion of interaction were summed up for each one of the five courses. Data was plotted in vertical bar charts according to the frequency of actions of the teacher and the student in each performance criteria. Everything is presented in one chart including the five different courses.

Two months after video recording the lessons, and after the students had already uploaded the institutional assessment of their teachers (this instrument was provided by the administration and is different from the questionnaires that were used in this study), the teacher performance assessment questionnaire was administered as well as the student performance self-assessment questionnaire. The time limit for the application was 15 to 20 minutes including both questionnaires. Previously, in the introductory session, the purpose of the questionnaire was explained. With the resulting data, averages were computed for each of the five criteria for didactic performance in each course, and they were organized in two independent tables: one just for teacher performance and the other for student performance.

Findings / Results

Criteria for Didactic Performance in the Observational Record

Figure 2 presents the results of the records obtained from the video of five lessons in five different high school science courses. The horizontal axis plots the indicators of performance from the teacher and from the students when they interact during a lesson. Frequency of appearance of the criteria for didactic performance is showed in the vertical axis.

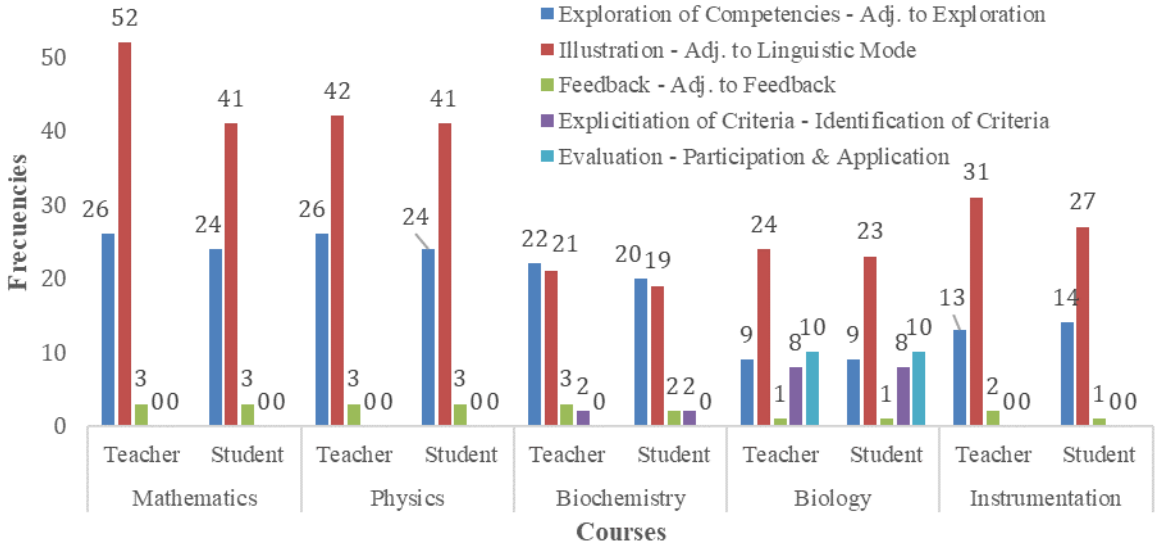

Figure 2 Record of Performance Actions in the Interactions Criteria by Course

Data from the study reflects that the didactic performance criterion with the highest frequency of occurrence is *Illustration – Adjustment to the linguistic mode and illustration*, presenting high proportions in all five courses. The *Exploration of competencies – Adjustment to the exploration of competencies* didactic performance criterion also presented high proportions in all five courses; in the Physics course, this criterion was positioned slightly higher than the Illustration criterion. The other criteria for didactic performance were less frequently presented.

Feedback – Participation and student Adjustment to Feedback was present in all five courses. The Explicitation of Criteria – Identification of Criteria didactic performance category was present in the Physics and Biochemistry courses, while the Evaluation – Evaluation and Application criteria was only present in the Biochemistry course.

Assessment of Teacher Performance and Self-Assessment of Student Performance

In Table 4 the results of the administration of the didactic performance questionnaires are presented considering the corresponding criteria for teacher performance and the criterion average in every course. It should be noted that the scale goes from 0 to 3. With that in mind, the greater the average punctuation described in the Table, the greater the frequency of occurrence that is reported.

Table 4. Descriptive Statistics of the Criteria for Teacher Performance in Every Course

	Exploration of competencies		**Explicitation of criteria**		**Illustration**		**Feedback**		**Evaluation**	
Course	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Mathematics	2.45	0.12	2.74	0.17	2.84	0.12	2.77	0.15	2.84	0.17
Physics	3.44	0.10	3.90	0.06	3.74	0.08	3.41	0.14	3.26	0.16
Biology	2.91	0.15	3.64	0.10	3.30	0.12	3.52	0.13	2.88	0.16
Biochemistry	2.96	0.15	3.39	0.10	3.27	0.14	3.18	0.15	2.89	0.18
Instrumentation	2.25	0.17	3.21	0.15	3.15	0.13	2.79	0.21	2.42	0.17

In Table 5 the results of student performance are presented as well as the average for each criterion in every course.

Table 5. Descriptive Statistics of the Criteria for Student Performance in Every Course

	Adjustment to the Exploration of Competencies		Identification of criteria		Adjustment to the Linguistic Mode and Illustration		Adjustment to Feedback		Evaluation – Participation and Application	
Course	**Mean**	**SE**	**Mean**	**SE**	**Mean**	**SE**	**Mean**	**SE**	**Mean**	**SE**
Mathematics	2.68	0.10	**3.29**	0.11	**3.24**	0.14	2.32	0.12	**3.26**	0.09
Physics	**3.00**	0.09	**3.48**	0.10	**3.54**	0.07	2.41	0.11	**3.48**	0.10
Biology	2.68	0.11	**3.24**	0.13	**3.32**	0.12	2.36	0.13	**3.36**	0.11
Biochemistry	2.82	0.12	**3.29**	0.10	**3.29**	0.10	2.39	0.12	**3.25**	0.10
Instrumentation	2.33	0.12	**3.13**	0.10	2.90	0.14	2.06	0.09	**3.21**	0.12

Discussion

The present study integrates the results of two strategies for the recollection of information regarding criteria for didactic performance in secondary level of education (High school) science courses. One strategy is based on the observational records of the classes and the other is based on a self-report questionnaire which was completed by each student regarding their teacher's performance and their own performance as students during the lessons that were videotaped for the observational records.

The first aspect that stands out is related to the results of the observational records. In every science class that was observed and analyzed the criterion of performance that occurred most frequently was the Illustration (teacher performance) – Adjustment to the linguistic mode and illustration (student performance) pair. The second didactic performance criterion that occurred most frequently was the Exploration of competencies – Adjustment to the exploration of competencies pair, and the third pair was the Feedback – Participation and Adjustment of the student to the feedback. Even though this last criterion pair was presented less frequently, it was still present in all five courses. On the contrary, the Explicitation of criteria – Identification of criteria pair and the Evaluation – Evaluation and application pair were scarcely presented and just made an appearance in one or two courses.

The data observed in this study associates, in part, with findings from other observational studies in secondary-level education classes. Peralta and Roselli (2015) observed college-level lessons in a Physics school and in a Psychology school and they found that didactic interaction of participation and guided interactions occurred most frequently in class, which are similar to the didactic performances of Illustration, Exploration of competencies, and Feedback – Participation. Likewise, Díaz et al. (2015), using observational records of college-level lessons in Spain and in Mexico, reported teacher feedback on student's questions, reinforcement of participation, and the encouragement of student participation as good practices, which in the case of this study can be related to the Feedback – Participation criterion that was the third most frequently presented. Furthermore, the data related to the Explicitation of criteria, Illustration, and Feedback, as well as the Exploration of competencies and Evaluation coincide with the teacher activities that Bell et al. (2019) highlight: classroom management, explanation of the topic, quality of the representation of the learning material (thematic content), learning assessment, and teaching strategies for learning and self-regulation to the students.

These results of the interactions in five different courses concur with the results reported by Velarde-Corrales and Bazán (2019) who used records of high school Math and Biology classes. Their results showed that the didactic performance criteria of Illustration – Adjustment to illustration and Exploration of competencies – Adjustment to the exploration were the most significant even though the Feedback – Adjustment to feedback criterion also was presented in both courses but in less proportion. Regarding the didactic performance of the teacher that was called Illustration, which was the less frequent, data in this study matched the findings of Borges and Falcón (2018) with their analysis of the function of a teacher in university-level classes. These authors found that the most frequent teacher performances were Linear explanations and Explanation with materials or resources.

Despite these findings, it is important to emphasize that the criteria for didactic performance related to Explicitation of criteria – Identification of criteria and Evaluation – Evaluation and application, were very low in occurrences and only appeared in two of the five courses. For future studies, it will be necessary to include more observations for each course in order to analyze the frequency of occurrence of these criteria to a greater extent. Both criteria of performance are related to the teaching/learning objective and to the process of formative evaluation. Considering the previous statement, these criteria provide the student with an assessment of what is learned and how to improve to achieve the learning objectives (Andersson et al., 2019; Krijgsman et al., 2019).

The second aspect to mention is the student's self-report regarding teacher performance and their own performance. Self-report data was above average, considering the 0 to 3 range, in every criterion and in all of the courses. For teacher performance, Explicitation of criteria, Illustration, and Feedback were the three criteria with high averages. For student performance, Compliance of criteria, Evaluation and Illustration were the three criteria with high averages. The coincidence between high ratings for Explicitation of criteria and Illustration (for teacher performance) and high self-

ratings for Adjustment to explicitation of criteria and Illustration (for student performance) stands out. Nonetheless, in all other criteria averages were moderately high (from 2.25 to 3.44 for teacher performance and from 2.06 to 3.00 for student performance).

These results relate to the findings regarding student assessment of teacher competencies and skills in higher-level education. Data from this study also relates to data reported in investigations that use college students' self-reports about teacher behavior during the practice of teaching (Grammatikopoulos et al., 2015; Nasser-Abu, 2017).

Although it is true that there is not a lot of investigation about student performance self-assessment in teaching/learning situations, the results of this study allow for better identification of the activities that are developed by the teacher and the student in the five criteria of performance when compared to just using the observational record analysis. Observational records also allow the identification of the performance criteria in didactic interactions, but many observations in different sessions are required for just one course. On the other hand, self-reports allow the tracing of general tendencies of behavior during different lessons of one course, this way, the student could be referring to various sessions of the same course in a semester. Furthermore, the results of the observational records and the self-report questionnaires, as well as the reliability and validity of both instruments, confirmed constructs for the criteria of teacher and student performance derived from the didactic performance model (Carpio et al., 1998; Irigoyen et al., 2011; Silva et al., 2014) that can identify and describe didactic interactions which allow for the improvement of teaching professional competencies and competencies for investigation (Galindo, et al., 2017).

Data From the Observational Records Versus Data from Self-Reports

Another aspect to discuss is the comparison between the data recollected from the observational record and the data from the didactic performance questionnaires. The first aspect to discuss is a certain correspondence that can be observed in the Illustration criteria according to the type of instrument and the method for analysis. Illustration was the didactic criteria that most frequently appeared in the observational records in all five courses. In regard to the self-reports, Illustration was the second didactic criterion with high average occurrences for teacher performance and the third criterion with high averages for student performance. In other words, independently of the instrument that was used to identify didactic interactions in the teaching of science courses at a secondary level, Illustration is the didactic performance criteria with higher occurrences.

The second aspect to discuss is related to the didactic performance criterion of Exploration of competencies which was the second most frequent criterion in the observation of all five science courses with moderate occurrence of teacher and student performance according to the self-report questionnaire. Similarly, the didactic performance criterion of Feedback appeared in medium proportion in the interactions record of all five courses. It was the third criterion with high averages in teacher performance that was assessed by the student questionnaire but in the student's self-assessment it was the lowest-ranked criterion even though it was above average.

Comparing these methodologies allowed the identification of two contradictory aspects; the first refers to the didactic performance criterion of Explicitation of Criteria and the other refers to the criterion of Evaluation. These criteria were presented in low proportions in the observational records but, both in the teacher performance assessment questionnaires and in the student self-assessment, Explicitation of criteria was the one that obtained higher averages while the Evaluation criterion was the second one with a high average in student performance.

These findings highlight the relevance of having more than one strategy to gather information regarding the didactic interaction and the relation between criteria for teacher performance and student performance. A self-report instrument with good construct validity can provide more information about the tendency of occurrence of these didactic performance criteria (from the teacher and the student) during a school year or a semester in different lessons of one course. This could also provide complementary information to the analysis of a lesson with an observational record system.

In this study in particular, the self-report questionnaire allowed the finding that, in the lessons of all five science courses, the didactic performance of Explicitation of criteria and didactic performance of Evaluation, according to the students, are frequently present (always or almost always) in the course lessons. This aspect could not have been known if only the data from the observational records was considered. In sum, the assessment of teacher performance with the questionnaires and student performance self-assessment in didactic interactions partially confirms the information that is obtained with observational records of didactic interactions in science courses of secondary-level education. Moreover, these instruments also provide information that is not identified in the observational records.

Conclusion

The analysis of teacher-student interactions is an important field of study because research discovering this matter can promote school excellence. In this research, teacher-student classroom interaction was evaluated using two types of data gathering methodologies (observation and verbal reports) to compare the frequency of occurrence of different aspects of teacher and student performance. The coincidences and differences found in both methodologies displayed

the social character of scientific investigation. On this subject, Callejo (2002) emphasized that every situation of observation is a social situation and that, because of this, the gathering of observer and observee is conditioned to its norms. In this case, the observation setting demands that its actors "act" based on dominant models, from general references, or from everyday activities that happen in their group of reference. Every observation implies the encounter of two worlds: the everyday life of the observed and the everyday life of the observer.

The transition between worlds is modulated by silence, which is conceived as a communicative element. "Silence becomes a symptom of the level of normative incorporation of the practices because the ones that reach a higher level of incorporation tend to be the ones that barely need to be spoken" (Callejo, 2002, p. 415). In this respect, silence is necessarily an active quietude since one is quiet because one knows. Callejo (2002) allows the characterization of different methodologies for the gathering of information. In this sense, observation can be characterized by the silence of the one who is quiet (observer) before the activity of someone else (observee) whereas with questionnaires, the one who is interviewed cannot be silenced. This explains the emergence of a greater number of categories of performance when using questionnaires in this study.

Callejo (2002) also recognizes that there would be a better understanding of the observed if silence could be managed and if norms were considered (the ones related to the observation, the references, and the practices); the norms that are the fundamentals of action for the observed and the ones for the observer. In short, the investigator must use different techniques for the recollection of data in order to get closer to the study of the experiences of the subjects without losing sight of the idea that not only the subject's actions are projected in the records, but that the records project the investigator's actions as well.

Recommendations

Observation of teacher performance in the classroom is a tool that school authorities and educational policy makers rely on to foster student learning, motivation, and teacher improvement. To become an effective evaluation tool, classroom observations may be a high-time and high-consuming resource option. For example, Hill and Grossman (2013) recommended that classroom observations should be subject-specific and content experts should participate in the process of rating teacher performance. Thus, the use of alternative data gathering methodologies should be implemented: verbal reports or discussion groups.

Callejo (2002) described discussion groups as a reunion of members from different social groups with the purpose of revealing the norm imbricated in a specific social context. In terms of silence, it alternates as members participate under the guidance of a moderator. Future research and school interventions could implement discussion groups to evaluate the impact of this social research methodology on teacher improvement. In this case, experts would moderate the interaction among teachers, students, and policy makers. One relevant topic to discuss is what good teaching looks like since Jones et al. (2022) have concluded that the use of universal teaching performance criteria may result in unfair teacher evaluations.

Finally, one of the main findings of the present study is that the use of student self-reports captures long-term educational practices. Therefore, future research could explore how well verbal reports correlate with aspects of teacher or student competencies in scenarios outside the classroom or that are required to promote learning transference. This analysis is important since the frequency and quality of teacher–student out-of-classroom interactions (e.g., office visits) have been linked to educational quality in college education (Hagenauer & Volet, 2014).

Limitations

Teacher – student interaction is a broad object of study. In the present research, emphasis was on competencies that can be unfolded in the classroom. A behavioral approach was favored considering data from verbal reports as complementary. Therefore, some limitations can be outlined.

First, the number of videotaped sessions was established arbitrarily. Future studies could consider different rationales such as filming the number of classes involved in a complete learning unit. This may reveal aspects of teacher and student performance that was overshadowed in the present study.

Secondly, competence development demands considering learning scenarios outside the classroom. As mentioned earlier, future research must consider filming teacher – student interaction in science labs, school projects, group work, academic events, or virtual interaction. These places and forms of relationships may shift traditional roles of power giving "voice" to students. Since silence is a key aspect (Callejo, 2002) to characterize and distingue among data gathering methodologies, it is important to broaden the analysis of didactic interaction beyond the classroom.

Thirdly, the results focused on frequencies of occurrences. Filmed material can be analyzed to withdraw findings regarding the quality of the didactic interaction and its role in learning the target scientific competencies in the observed science classes. For example, Durksen et al. (2017) found that teacher performance such as eye contact or individual feedback is appreciated by math students. This kind of analysis was beyond the present research objective, but they could be addressed in future studies.

Authorship Contribution Statement

Bazán-Ramírez: Conceptualization, design, data analysis / interpretation, writing. Velarde-Corrales: Concept and design, data acquisition, statistical analysis, supervision. Rodríguez-Pérez: Analysis, Editing/reviewing, critical revision of the manuscript, writing. Guerrero-Barrios: Drafting the manuscript, Analysis, technical or material support. Anaya-González: Supervision, critical revision of the manuscript, technical or material support.

References

Andersson, U. B., Löfgren, H., & Gustafson, S. (2019). Forward-looking assessments that support students' learning: A comparative analysis of two approaches. *Studies in Educational Evaluation, 60*, 109-116. https://doi.org/10.1016/j.stueduc.2018.12.003

Anguera, M. T., Blanco, Á., Hernández, A., & Losada, J. L. (2011). Diseños observacionales: Ajuste y aplicación en psicología del deporte [Observational designs: their suitability and application in sports psychology]. *Cuadernos de Psicología del Deporte, 11*(2), 63-76. https://revistas.um.es/cpd/article/view/133241

Bazán-Ramírez, A., Capa-Luque, W., Bello-Vidal, C., & Quispe-Morales, R. (2022). Influence of teaching and the teacher's feedback perceived on the didactic performance of Peruvian postgraduate students attending virtual classes during the COVID19 pandemic. *Frontiers in Education*, *7*, 1-16. https://doi.org/10.3389/feduc.2022.818209

Bazán-Ramírez, A., Martínez, X., & Trejo, M. (2009). Análisis de interacciones en clases de español de primer grado de primaria [Analysis of interactions in first grader's Spanish classes]. *Interamerican Journal of Psychology/Revista Interamericana de Psicología, 43*(3), 466-478. http://pepsic.bvsalud.org/pdf/rip/v43n3/v43n3a06.pdf

Bazán-Ramírez, A., Pérez-Morán, J. C., & Bernal-Baldenebro, B. (2021). Criteria for teaching performance in psychology: Invariance according to age, sex, and academic stage of Peruvian students. *Frontiers in Psychology*, *12*, 1-13. https://doi.org/10.3389/fpsyg.2021.764081

Bazán-Ramírez, A., & Velarde-Corrales, N. (2021). Autoreporte del estudiantado en criterios de desempeño didáctico en clases de psicología [Students self-report within didactic performances criteria in psychology classes]. *Journal of Behavior, Health and Social Issues*, *13*(1), 22-35. https://bit.ly/3uedxJt

Bell, C. A., Dobbelaer, M. J., Klette, K., & Visscher, A. (2019). Qualities of classroom observation systems. *School Effectiveness and School Improvement*, *30*(1), 3-29. https://doi.org/10.1080/09243453.2018.1539014

Borges, A., & Falcón, C. (2018). Protocolo de observación de la función de explicación (PROFE): Un instrumento para operacionalizar la transmisión de contenidos por parte del profesorado [Observation protocol of the explanation function (OPEF): An instrument to operationalize the transmission of contents by teachers]. *Universitas Psychologica*, *17*(3), 1-12. https://doi.org/10.11144/Javeriana.upsy17-3.pofe

Borges, A., Falcón, C., & Díaz, M. (2016). Creation of an observational instrument to operationalize the transmission of contents by university teachers. *International Journal of Social Science Studies*, *4*(7), 82-89. https://doi.org/10.11114/ijsss.v4i7.1596

Callejo, J. (2002). Observación, entrevista y grupo de discusión: El silencio de tres prácticas de investigación [Observation, interview and discussion group: The silence of three research practices]. *Revista Española de Salud Pública*, *76*(5), 409-422. https://doi.org/10.1590/S1135-57272002000500004

Carpio, C., Pacheco, V., Canales, C., & Flores, C. (1998). Comportamiento inteligente y juegos de lenguaje en la enseñanza de la psicología [Intelligent behavior and language games in the teaching of psychology]. *Acta Comportamentalia*, *6*(1), 47-60. https://bit.ly/3O6FcE2

Chan, W. M. (2018). Teaching in higher education: Students' perceptions of effective teaching and good teachers. *Social Sciences and Education Research Review*, *5*(1), 40-58. https://bit.ly/3zHbTDy

Copur, Y., & Thacker, I. (2021). A comparison of perceived and observed learning from professional development: Relationships among self-reports, direct assessments, and teacher characteristics. *Journal of Teacher Education*, *72*(2), 138-151. https://doi.org/10.1177/0022487119899101

Corral-Verdugo, V. (1997). Dual 'realities' of conservation behavior: Self-reports vs observations of re-use and recycling behavior. *Journal of Environmental Psychology*, *17*(2), 135-145. https://doi.org/10.1006/jevp.1997.0048

Díaz, M. C., Borges, A., Valadez, M. D., & Zambrano, R. (2015). Valoración de buenas prácticas docentes a través de observación sistemática [Assessing of good teaching practices through systematic observation]. *Universitas Psychologica*, *14*(3), 913-922. https://doi.org/10.11144/Javeriana.upsy14-3.vbpd

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, *44*(4), 237-251. https://doi.org/10.3102/0013189X15584327

Durksen, T. L., Way, J., Bobis, J., Anderson, J., Skilling, K., & Martin, A. J. (2017). Motivation and engagement in mathematics: A qualitative framework for teacher-student interactions. *Mathematics Education Research Journal*, *29*(2), 163–181. https://doi.org/10.1007/s13394-017-0199-1

Galindo, L., Silva, H., Serrano, V., Rocha, E., & Galguera, R. (2017). Aprendizaje por observación de interacciones didácticas de ilustración y retroalimentación [Interbehavioral survey about learning by observation in didactic interactions of illustration and feedback]. *Interacciones*, *3*(3), 131-140. https://doi.org/10.24016/2017.v3n3.71

García, I., Romero, S., Rubio, S., Flores, V. J., & Martínez, A. (2015). Comparación de prácticas inclusivas de docentes de servicios de educación especial y regular en México [Comparison of inclusive practices of regular and special education services teachers in Mexico]. *Actualidades Investigativas en Educación*, *15*(3), 238-254. https://doi.org/10.15517/aie.v15i3.20671

Gitomer, D. H. (2019). Evaluating instructional quality. *School Effectiveness and School Improvement*, *30*(1), 68-78. https://doi.org/10.1080/09243453.2018.1539016

Grammatikopoulos, V., Linardakis, M., Gregoriadis, A., & Oikonomidis, V. (2015). Assessing the students' evaluations of educational quality (SEEQ) questionnaire in Greek higher education. *Higher Education*, *70*(3), 395-408. https://doi.org/10.1007/s10734-014-9837-7

Hagenauer, G., & Volet, S. E. (2014). Teacher–student relationship at university: An important yet under-researched field. *Oxford Review of Education*, *40*(3), 370-388. https://doi.org/10.1080/03054985.2014.921613

Hill, H., & Grossman, P. (2013). Learning from teacher observations: Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review*, *83*(2), 371–384. https://doi.org/10.17763/haer.83.2.d11511403715u376

Hugh-Pennie, A. K., Park, H. S. L., Luke, N., & Lee, G. T. (2018). Applied behavior analysis as a teaching technology. In V. C. Bryan, A. T. Musgrove & J. R. Powers (Eds.), *Handbook of research on human development in the digital age* (pp. 330-362). IGI Global. https://doi.org/10.4018/978-1-5225-2838-8.ch015

Ibáñez, C. (1999). Conducta de estudio: El papel de identificar criterios en el discurso didáctico [Study behavior: The role of identifying criteria in didactic speech]. *Acta Comportamentalia*, *7*(1), 47-66. https://bit.ly/3tJMPrA

Ibáñez, C. (2007). Un análisis crítico del modelo del triángulo pedagógico. Una propuesta alternativa [A critical analysis of the pedagogic triangle model. An alternative proposal]. *Revista Mexicana de Investigación Educativa*, *12*(32), 435-456. https://bit.ly/3QEZCpk

Irigoyen, J., Acuña, K., & Jiménez, M. (2011). Interacciones didácticas en educación superior. Algunas consideraciones sobre la evaluación de desempeño [Didactic interactions in higher education. Some considerations about performance evaluation]. In J. Irigoyen, K. Acuña, and M. Jiménez (Eds.), *Evaluación de desempeños académicos* (pp. 73–96), Universidad de Sonora. https://bit.ly/3n3R79r

Jones, N. D., Bell, C. A., Brownell, M., Qi, Y., Peyton, D., Pua, D., Fowler, M., & Holtzman, S. (2022). Using classroom observations in the evaluation of special education teachers. *Educational Evaluation and Policy Analysis*. Advance online publication. https://doi.org/10.3102/01623737211068523

Joyce, J., Gitomer, D. H., & Iaconangelo, C. J. (2018). Classroom assignments as measures of teaching quality. *Learning and Instruction*, *54*, 48-61. https://doi.org/10.1016/j.learninstruc.2017.08.001

Krijgsman, C., Mainhard, T., van Tartwijk, J., Borghouts, L., Vansteenkiste, M., Aelterman, N., & Haerens, L. (2019). Where to go and how to get there: Goal clarification, process feedback and students' need satisfaction and frustration from lesson to lesson. *Learning and Instruction*, *61*, 1-11. https://doi.org/10.1016/j.learninstruc.2018.12.005

Liu, J., & Cohen, J. (2021). Measuring teaching practices at scale: A novel application of text-as-data methods. *Educational Evaluation and Policy Analysis*, *43*(4), 587–614. https://doi.org/10.3102/01623737211009267

Mares, M. G., Rueda, E., Rocha, H., Rivas, O., González, L. F., & Carrascoza, C. A. (2020). Efectos del grado de especificación del criterio de logro sobre la conducta docente y el aprendizaje escolar [Effects of achievement-criteria specification degree on teaching behavior and school learning]. *Revista Mexicana de Análisis de la Conducta*, *46*(1), 148-178. http://doi.org/10.5514/rmac.v46.i1.76952

Martínez, R., & Moreno, R. (2002). Integración de teoría sustantiva, diseño de pruebas y modelos de análisis en la medición psicológica [Integration of substantive theory, test design and models of analysis in measurements of the psychological]. In A. Bazán & A. Arce (Eds.), *Estrategias de evaluación y medición del comportamiento en Psicología* (pp. 87–119). ITSON-UADY.

Mislevy, R. J. (1993). Foundations of a new test theory. In N. Frederiksen, R. J. Meslevy, & I. I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 19–39). Lawrence Erlbaum.

Morales, G., Peña, B., Hernández, A., & Carpio, C. (2017). Competencias didácticas y competencias de estudio: Su integración funcional en el aprendizaje de una disciplina [Didactic competencies and competencies for studying: Their functional integration in learning a discipline]. *Alternativas en Psicología*, *37*(1), 24-35. https://bit.ly/3QtHr5W

Nasser-Abu, F. (2017). Teaching in higher education: Good teaching through students' lens. *Studies in Educational Evaluation*, *54*, 4-12. https://doi.org/10.1016/j.stueduc.2016.10.006

Peralta, N. S., & Roselli, N. D. (2015). Los sistemas de interacción generados por la impronta didáctica del docente [The interaction systems generated by the teacher's didactic imprinting]. *Propósitos y Representaciones*, *3*(2), 155-177. https://doi.org/10.20511/pyr2015.v3n2.85

Ribes, E. (2008). Educación básica, desarrollo psicológico y planeación de competencias [Basic education, psychological development and competency planning]. *Revista Mexicana de Psicología*, *25*(2), 193-207. https://bit.ly/2Ubzmoi

Ruiz, E., Cruz, J. L., & Hernández, T. F. (2020). Análisis secuencial del discurso docente ante respuestas del alumno en segunda lengua [Sequential analysis of the teaching discourse upon student's response in second language adquisition]. *Revista Electrónica de Psicología Iztacala*, *23*(2), 724–750. https://bit.ly/3HAVBhe

Salinas, M. I. (2017). Gestión de la evaluación del desempeño docente en aulas virtuales de un proyecto de blended-learning [Management of teacher assessment in virtual classrooms of a blended-learning project]. *Ciencia, Docencia y Tecnología*, *28*(54), 100-129. https://bit.ly/39wuIyy

Scherer, R., Nilsen, T., & Jansen, M. (2016). Evaluating individual students' perceptions of instructional quality: An investigation of their factor structure, measurement invariance, and relations to educational outcomes. *Frontiers in Psychology*, *7*, 110. https://doi.org/10.3389/fpsyg.2016.00110

Silva, H., Morales, G., Pacheco, V., Camacho, A., Garduño, H., & Carpio, C. (2014). Didáctica como conducta: Una propuesta para la descripción de las habilidades de enseñanza [Didactic as behavior: A proposal for the description of teaching skills]. *Revista Mexicana de Análisis de la Conducta*, *40*(3), 32–46. https://doi.org/10.5514/rmac.v40.i3.63679

Simpson, Z., & Bester, J. (2017). Cognitive demand and student achievement in concrete technology study. *Journal of Professional Issues in Engineering Education and Practice*, *143*(2), Article 04016022-1-8. https://doi.org/10.1061/(ASCE)EI.1943-5541.0000307

Üstünlüoglu, E., & Güngör-Culha, D. (2012). Investigating student evaluation of teachers by using latent class analysis: A case study at a tertiary level. *International Journal of Education*, *4*(3), 147-159. https://doi.org/10.5296/ije.v4i3.1811

Van der Lans, R. M., van de Grift, W., & van Veen, K. (2018). Developing an instrument for teacher feedback: Using the Rasch model to explore teachers' development of effective teaching strategies and behaviors. *The Journal of Experimental Education*, *86*(2), 247-264. https://doi.org/10.1080/00220973.2016.1268086

Vargas, J. S. (2020). *Behavior analysis for effective teaching* (3rd ed.). Routledge. https://doi.org/10.4324/9780429442575

Velarde-Corrales, N., & Bazán, A. (2019). Sistema observacional para analizar interacciones didácticas en clases de ciencias en bachillerato [Observational system to analyze didactic interactions in science classes in bachelor]. *Revista de Investigación en Psicología*, *22*(2), 197-216. https://doi.org/10.15381/rinvp.v22i2.16806