UNIVERSIDAD DE CHILE
FACULTAD DE FILOSOFÍA Y HUMANIDADES
DEPARTAMENTO DE LINGÜÍSTICA

THE INTERDEPENDENCY BETWEEN INSTITUTIONAL RESOURCES, TEACHERS' FORMATION, AND THE VALIDITY AND RELIABILITY OF ENGLISH L2 RUBRICS IN CHILE: A CASE STUDY

MARIANA CÓRDOVA
LAURA DAVICO
JAVIERA GUZMÁN
LUCAS MEDINA
BRYAN MORA
TAMAR MOORE
LINMARA OYARZÚN
BENJAMÍN SILVA

Informe Final de Seminario de Grado para optar al grado de Licenciado en Lengua, Lingüística y Literatura Inglesas

Profesor guía: Daniel Muñoz Acevedo

SANTIAGO DE CHILE
DICIEMBRE 2019

ACKNOWLEDGMENTS

Agradezco de corazón a mi equipo de tesis. Tuve la suerte de trabajar con un grupo de ñoños inteligentes y motivados, comprometidos igual que yo con poner nuestro granito de arena para mejorar la educación en Chile. Cabrxs, estoy orgullosa de ustedes por su pasión por la pedagogía, y sé que serán grandes profesionales, porque son grandes personas. Los felicito por la fortaleza que tuvieron este año para trabajar tan arduamente en la tesis a la vez que lidiaban con dificultades y obstáculos en su vida personal. Admiro mucho a cada uno de ustedes, y no me cabe duda de que serán capaces de cumplir todas sus metas. Por último, gracias por aguantarme llegar tarde a las reuniones, no se cómo no me mataron. Tal vez es por nuestra hermosa filosofía común: efectivamente, tiene que asepta.

Agradezco al nuestro profe guía, Daniel. Gracias a usted y a su pasión por la evaluación fuimos capaces de crear una tesis que nos dejó a todos contentos y orgullosos. Paso a paso y con su ayuda logramos construir algo bacán. De verdad es un excelente profesor guía. Pero no todo es trabajo, así que también agradezco todas esas conversaciones sobre la vida y sus vicisitudes. Es una persona extremadamente interesante, y siento que aprendí mucho de usted.

A las profes Ximena Tabilo y Rosa Bahamondes, les agradezco haberme inculcado su amor por la lingüística aplicada. Son excelentes docentes y estoy muy feliz de haber sido su alumna. Gracias a sus enseñanzas ahora sé que puedo usar la lingüística para lograr metas que he tenido desde hace años. Gracias a ustedes podré construir mi futuro con las herramientas de base que necesitaba.

A mi alma gemela académica, Luli, te agradezco por los infinitos trabajos y presentaciones que hemos hecho juntas. Nunca en mi vida había estado tan en sintonía con alguien, siento que nos entendemos perfectamente. Hablo en presente porque espero que sigamos así. After all, todavía nos queda un año en la licenciatura y, quién sabe, tal vez entremos juntas a la pedagogía. Luli, sobre todo te agradezco por haber tomado conmigo la difícil decisión de cambiarnos de la Cato a la Chile. Estoy segura de que fue lo mejor que

podríamos haber hecho. Aunque ese año fue tormentoso, nos apañamos siempre en todo, y en realidad es lo que hemos hecho desde el momento en que nos hicimos amigas. Hemos sufrido mucho, pero también hemos sabido salir adelante juntas. Agradezco mucho haberme topado contigo en la vida.

Agradezco también a mi familia, que siempre ha apoyado todos mis sueños, ha creído en mi potencial, y me ha aguantado en mis momentos de estrés y mal humor. Agradezco en especial a mi mamá. Tú fuiste mi primera profesora, y al ver tu entrega por tus alumnos despertaste en mí la pasión por enseñar. Sé que es un trabajo demandante y duro, y por eso te admiro mucho y te agradezco por haber apoyado mi decisión. Espero tener la misma fortaleza que tú en mi camino profesional y personal. Si algún día logro ser un aporte para la educación en Chile, espero que sepas que fue gracias a ti.

Finalmente, agradezco a mi hijo-perro, Dagoberto, aunque no entienda nada y aunque al profe le parezca ridículo. Eres mi compañero peludo y te amo más de lo que la gente puede comprender. El lenguaje común que hemos creado nos une y nos hace eternos. Me acompañas en mis penas y mis alegrías, y siempre me das todo tu amor. Llegaste a darle luz a mi vida. Como la etimología de tu nombre lo dice, eres aquel que brilla como el sol.

Mariana Córdova

Quiero partir agradeciendo *al grupo de seminario* por ser tremendas personas, inteligentes y bacanes. Su buena onda y dedicación hizo de este trabajo uno mucho más ameno y lleno de aprendizajes. También quiero agradecer *al profesor Daniel Muñoz* por tirarnos pa' arriba todo el rato, y sobretodo por su simpatía que nos hizo sentir que era uno más del grupo. *A la Universidad de Chile*, por la vocación social y pedagógica que me inculcó, y la conciencia que brotó en mí a lo largo de este camino. En cuanto a mi amor a la educación, quiero agradecer *a la profesora Ximena Tabilo* por llenarme de amor hacia la lingüística aplicada, y *a la profesora Rosa Bahamondes* por hacerme reafirmar mi amor por la pedagogía crítica e inclusiva. No quisiera dejar afuera a muchxs otrxs profesorxs, que estuvieron ahí apoyando y enriqueciendo todo mi aprendizaje de una manera crítica, humana, y cariñosa.

Por otro lado, quiero agradecer a *mi familia*, en especial a *mis papás*, por haberme apoyado siempre con lo que quería hacer y darme la libertad de seguir mi camino elegido, haciéndome sentir que me la podía. También *a mis hermanos*, que son un ejemplo de perseverancia y autenticidad en sus carreras. Mención especial *a la Mantra*, mi hija gatuna, por hacerme feliz día a día e inyectarme ese amor y esas ganas de vivir con solo mirarla. Quisiera agradecer además *a mis abuelos Humberto Giannini y Luisa Eguiluz* y todo su legado, por ese espíritu literario, social, y consciente que fui heredando mientras crecí.

Por último quisiera agradecer *a mis amigxs* en general, quienes me acompañaron siempre a lo largo de estos años y me hicieron compañía. En especial quisiera agradecer *a la Mari Córdova*, coautora de esta tesis, por ser mi fiel compañera de universidad, de trabajos, y de ambiciones, con la que siempre me sentí en sintonía desde el primer año en la PUC hasta el día de hoy. *A mi eterno compañero de vida y mejor amigo, el Lalo*, por no dejarme nunca y haber crecido todos estos años conmigo. Por último, *al Tebi, mi pololo,* una lucecita que llegó a mi vida a mitad de año, volviéndola más linda y haciendo de este proceso uno mucho más llevadero con su infinito amor y apañe. Gracias *a toda la gente que me acompañó estos años*, estoy muy feliz y orgullosa de todo este camino

Laura Davico Giannini

En primer lugar, quiero agradecerle a Dios por haberme puesto en los lugares y momentos indicados, los cuales eventualmente me llevarían hasta donde esto, gracias maestro por tu gran sabiduría y por querer siempre lo mejor para mí. Gracias a mis amados padres Mabel y Marcos por enseñarme el arte de ser feliz, por motivarme siempre cuando el camino se hizo complicado, por haber confiado siempre en las decisiones que fui tomando, esto definitivamente es para ustedes. También quiero agradecer a mi hermano José Ignacio (mi cote) *because you are my person*, gracias por regalarme tu amor infinito, gracias por tu música, tu cuidado y tu arte que me animan a seguir, gracias por tu verdad y gracias por haber sido el centinela que me llevó a escoger el idioma como mi futuro: *hasta que nos riámos en las estrellas.* Quiero agradecer también a mi hermana mayor, mi madrina Pía, gracias por estar siempre a mi lado, por retarme cuando me estresaba, por hacerme onces ricas cuando estuve cansada, por estar siempre pendiente de las cosas que necesitaba para la Universidad, por siempre creer en mí y confiar en que llegaría hasta este momento, gracias por tu amor incondicional.

En segundo lugar, quiero agradecer a mi equipo de seminario y al profe Daniel por el constante apañe cuando las cosas se pusieron difíciles, por la motivación que tuvimos para lograr esto que ya está en nuestras manos y por haber compartido nuestros sueños y metas a lo largo de este año. Quiero agradecer a aquellas que estuvieron a lo largo de estos cuatro años en mi camino, María Jesús gracias por las conversaciones infinitas, los abrazos y los llantos compartidos, por compartir tus conocimientos conmigo y por el apañe que superó cualquier distancia. Yihad, gracias por ser una persona completamente verdadera, por todos esos trabajos de lite y que hoy podemos decir que lo logramos, eres una persona increíble y estoy muy agradecida de haberte encontrado. Anto, amiga, gracias por tu tremenda humildad, por enseñarme tanto a lo largo de este último tiempo, por tu compresión, tus retos, los miles de cigarros compartidos y sobre todo por quererme y apoyarme en los momentos difíciles. Javi, gracias por ser ese pilar en donde puedo descansar espiritualmente, por toda tu oración, por tus buenos deseos cada vez que los necesité, por tu confianza infinita y el amor incondicional que irradias. Finalmente, Cami,

gracias por hablarme con claridad cuando las cosas parecían oscuras, gracias por las risas compartidas, por tu frontalidad y transparencia y por tu amor incondicional.

También quiero agradecer a mis amigos de la vida, Paloma por estar desde hace 15 años compartiendo mis sueños y metas. Mauricio por ser aquel Cristo vivo en mi vida. Vale por ser la voz de mi consciencia y recordarme siempre quién soy. Mili, gracias por aquella unión que tenemos, que el cordón nunca se corte para que puedas abrazarme y aconsejarme antes de si quiera saber que lo necesito. Cynthia, gracias por creer que podría lograr cualquier cosa que me propusiera, por todo el ánimo que me entregaste durante el año y por no dejar que tirara la toalla jamás. A mis Matonas, por haber estado conmigo desde el principio, por haber hecho del estrés algo colectivo en donde nos dimos animo entre todas, por ser un mandala humano que solo aporta protección y buenas energías, gracias infinitas.

Quiero agradecer a mis profesores y profesoras que siempre estuvieron ahí pendientes y preocupados de nuestra salud mental, especialmente a la profesora Rosa Bahamondes y Pascuala Infante, gracias infinitas por ser grandes referentes para mi carrera pedagógica y por enseñarme a nunca perder la humanidad ni la pasión en lo que hago.

Gracias Patricio por haberme acompañado durante estos 4 años de carrera, por creer en mí, por haberme acompañado en los momentos difíciles de esta carrera, por las tardes de estudio, las clases de Excel, por enseñarme a ser cada vez mejor y por nunca dudar de que llegaría a este momento, muchas gracias por confiar en mí.

Finalmente, gracias a mi familia entera, a mis ancestros, Humilde, Hortensia, Luciana (nury), sé que desde donde estén celebran conmigo. De manera muy especial quiero agradecerle a mi tío Julio, gracias por los cigarros compartidos, por la sabiduría que me pudiste entregar, por nunca perder tu humor, sé que me acompañarás siempre, desde donde estés.

Javiera Guzmán

Primero que nada debo agradecer más que a nadie a mi grupo de seminario y al profesor guía, Daniel Muñoz, sin los cuales nada de esta tesis podría haber sido posible. Claramente hubo momentos difíciles pero aquí estamos, y lo que importa es que finalmente logramos terminar con la tortura y nos fue como nos tenía que ir.

Quisiera agradecer también a mi familia, a mis tatas que ya no están, y a mi madre y tía (y cómo no, a mi peluche favorito, Gusito), quienes aunque con el poco conocimiento de lingüística que tienen siempre se dieron el tiempo de escuchar cada una de mis teorías y conclusiones locas. Sin duda las muchas veces en donde yo les contaba sobre la tesis a la hora de la once me sirvieron mucho para poder lograr terminarla y seguir vivo en el intento. Realmente sirve mucho hablar con gente que no tiene idea de lo que les estás hablando, es básicamente un monólogo que al fin y al cabo te ayuda a salir del hoyo. En su ignorancia, sus constantes preguntas siempre sirvieron para aclarar mis dudas.

Infinitas gracias a mis profesores del colegio, sin quienes no hubiese descubierto mi amor por la lengua inglesa. A la Miss Karla, quien se ganó toda mi atención en su clase el día que me echó de la sala por pedir un sacapunta. Cómo no iba a terminar amando el inglés si sus argumentos eran tan re-buenos. También debo agradecer a Mr Hans, quien no solo fue mi profesor, sino también casi un hermano mayor. Gracias por ser mi mentor y por enseñarme fonética, por obsesionarme con las transcripciones, que por cierto mucho me sirvieron en mis años de carrera. Aunque aún me debas $50.000 (sin contar el interés), te estimo un montón; eres lejos una de mis mayores influencias en la vida.

También a mis profesores de la universidad, quienes me impartieron su conocimiento y me hicieron el monstruo de la lingüística que soy hoy en día. Muchas, pero muchas gracias por enseñarme todo lo que sé. Espero nunca decepcionarles.

A mis queridas amigas, Sara y Yihad, las que nunca faltaron cuando quería desahogarme. Realmente el pelambre con ustedes es 10/10 y lo recomiendo siempre. Aunque ustedes estaban en las mismas con sus propias tesis nunca nos faltó tiempo para darnos un ratito de descanso para simplemente hablar y soltar una más que merecida risotada. Espero que mis ocurrencias y desventuras les hayan servido para no querer morirse tanto.

Quiero agradecer, finalmente, no solo a Dios, sino también a la música que me acompañó en este camino, en especial a dos de mis cantantes favoritos, Hozier y Halsey, mejor conocidos (por mí) como el único hombre hétero y cisgénero que amo y mi polola #4. Honestamente sin ellos mi vida sería distinta. Con su música ellos me hacen querer seguir viviendo, ya que me dan esperanzas de que aún existe gente que vale la pena seguir como ídolos. Muchas gracias por existir y por ser tan hermosas personas.

Lucas Medina Vega

Gracias a *mis padres y mis hermanos* por haber sido un pilar fundamental en el tiempo que estuve en la u, y también en mi vida. Gracias por haberme enseñado todo lo que me han enseñado, gracias por haber desaprendido algunas cosas también, lo valoro demasiado. También gracias por siempre apoyarme a pesar de todo, son los mejores<3. Los amo.

Gracias a *mis compañeros de seminario* por todo el esfuerzo que dedicaron para que la tesis quedara bacán.

Gracias al *profe Daniel* no solo por guiarnos y apoyarnos en este proceso sino que también por ser tan pulento y buena persona. He aprendido mucho de usted.

Gracias a mis amigos *Coke, Mati, Blazz y Nico*. Son bacanes y he aprendido demasiado de ustedes. Gracias por todos los consejos y las experiencias que hemos vivido juntos, las atesoro en mi kokoro uwu. También gracias a la *Paty* por todos esos cotorreos, se pasó bn.

Gracias a la *profe Pascuala* por ser la mejor. Le agradezco cada consejo y palabra de aliento que recibí. Es maravillosa.

Gracias al Gnar por ser el mejor gato. Gracias por todos los ronroneos, las amasadas, los rasguños y el cariño. Gracias por ser tan perceptivo y consolarnos en los momentos sad. Contigo he aprendido algunas cosas sobre el amor que no pide nada a cambio, eres bacan (aunque tienes tus momentos a lo *malDitto*).

Finalmente, garsias a mi compañera de vida, *Lin*. Eres la mejor persona que he conocido y te agradezco por todo. Garsias por ser como eres. He aprendido un montón de ti en tantos aspectos que costaría nombrarlos en este momento. Garsias por tu amor, los memes y las veces que me has ayudado. Espero haber retribuido aunque sea un poco de lo que tu me has dado. Te amo. Tiene k asepta…

Bryan Mora

En vista el cierre del largo proceso que ha sido la tesis, doy un espacio para agradecer a quienes me han apoyado durante este tiempo. A mis padres, *ñi poyen*, familia, amigos, profesores y grupo de seminario.

Amulepw taiñ weychan

Tamar Moore Urzúa

Lucas, gracias por ser la persona que me ha motivado a seguir haciendo esto más allá de todo. En los días malos, eres un rayo de esperanza que me dispara por la espalda. Has sido el gemelo que contribuye a mi visión del mundo y de lo que es la vida. Este año fue particularmente difícil, y debo decir que me he dado cuenta lo que más atesoro de mi familia eres tú. Te amo mas que a nadie, gracias por apoyarme, por ser tan buen hermano, gracias por existir.

Gracias al equipo de la tesis por siempre seguir y encontrar la forma de lograr las cosas.

Grax profe por todas esas peleas chistosas, las buenas discusiones y el apoyo que nos ha dado en todo este proceso.

Gracias al lol por darme tan buenos momentos y ayudarme a conocer gente bacán, a desestresarme, y a inspirar mi arte.

Gracias a los memes que construyen lo que digo cada día, que me hacen reír y a ratos forman parte de mi identidad.

Bryan, gracias por estar ahí, por ser mi amigo, mi confidente, mi compañero, mi pareja, y todo lo que una persona puede pedir. Ya van a ser tres años desde que esto comenzó, y estoy tan feliz de que podamos crecer y aprender juntos, de que hayamos estado juntos en todo este proceso en la u, y de que podamos recorrer nuestro camino dándonos la mano. Gracias por aguantar mi flojera, y siempre sacar lo mejor de mi. Te amo gordito. Pd: si lees esto, me debes una Vg burger.

Gracias a mis papas por el apoyo que me han dado en estos años, y por enseñarme a no ser como ellos. Igual los quiero.

A mis compitas muchas gracias por ser tan bacanes y compartir su vida conmigo, por todos los tarreos, las juntas, las conversaciones hasta tarde y el cariño; Mati, Nico, Blazz, Paula, Negra, Patita y Coke.

Gnar, Venus y Neptuno, gracias por ser tan geniales, por recibirme con cariño, por acurrucarse conmigo, por langüetearme y dejarme llena de pelos, por rajuñarme y morderme, por no haberme olvidado cuando me fui por tanto tiempo.

Aweli, gracias por ser tan buena con nosotros. Usted fue crucial este año en todo lo que sucedió, gracias por su apoyo y sus sabios consejos.

Fue un año de shit, pero como dijo un sabio, tiene k asepta.

Linmara Oyarzún

Como persona y estudiante de esta carrera, quisiera primero partir por agradecer a mi familia, la incondicional. Sin su apoyo quizás no sería nadie ni la persona que soy, ni la que voy a llegar a ser algún día. Quería agradecer a mi mama por siempre estar apoyándome en todo sentido, por ser la persona más importante que tengo, por sacarse la cresta todos los días por mí y mi hermana. Su formación y educación lo han sido todo para mí. Ojalá fueras eterna viejita, es lo único que pido. También quería agradecer a mi abuela por cumplir el mismo rol que mi madre, por siempre estar apoyándome en todo sentido y tirándome para arriba. Al igual que mi mama, quiero que seas eternas y que siempre estés conmigo en todo buen y mal momento. Agradezco también a mi papá por estar siempre apoyando y subiéndome el animo cuando estaba bajo, o aportando algún consejo bueno de la vida o sobre cocina, "es un buen tipo mi viejo", también se eterno, viejo. No podía finalizar esto sin agradecer a la Panchita, hermana mía de mí, mis días no serían lo mismo sin ti en el depa, siempre serás mi otra mitad, espero que te vaya bien en todo lo que viene por delante para ti.

Siendo también de la familia, agradezco a mis tíos de La Serena, a mis "pairinos", por siempre brindar su apoyo en todo momento, haya sido bueno o malo. Mencionar también a mis tios de Santiago, por apoyarme en todo momento, por recibirme en su hogar por un tiempo. Como no agradecer a mis primitos lindos, que son prácticamente mis hermanos, el Alfoncho, la Feñi, los gemes Pancho y JP, la Valita, y la Yoshi Rosario.

Agradezco también a mis compañeros y amigos de este seminario, nos sacamos la cresta para esto cabros y nos va a salir la raja, gracias por ser un grupo super apañe cuando el otro lo necesito, y por trabajar tan bacán. También, a mis compitas de Ovalle, mi ciudad natal, al Chupao, al Janichi, al Lea qqq, al Yean, al Keko, al Juaco. Los primeros pasos los di junto a ustedes y gracias por cada momento, incluso después de haber terminado el colegio de monjas. Obviamente agradecer también a quienes siempre estuvieron junto a mi en este proceso universitario, los ctm que, los peaky flaiters, los mismísimos, señoras y señores, sin más preámbulos, formación original. Al Pasto (yo te dije wn), al Poto (degenerado), al Aguirresamba, a la Danita Gonzales, al Mikel, a la Luli Dabrincci y la Mari. Sin ustedes no se que habría hecho estos cuatro años universitarios, probablemente

algo productivo. Sus consejos fueron lo mejor, al igual que su apoyo y los momentos que compartimos junto, que obviamente no olvidaré nunca.

Quería también mencionar a los profesores que mejoraron a más no poder mi experiencia universitaria. A lxs profes Giorgos, Maria Antonieta, Rosa, Pascuala, Novoa, Pichihuenche, Bonfanti, Ortiz, Carmencilla, Ruben, y muchos más que guiaron mi camino, siendo un pilar fundamental en la educación de mi vida. Por supuesto agradezco al profe Daniel Muñoz, por guiar este trabajo, aconsejarnos, darnos comida, y por aguantarnos como grupo.

Agradezco de igual forma a mis gringos queridos, Makiah Sullenger, Amber Hoffman, y la familia Sullenger, Ashley, Angela, Hannah y Trin.. Gracias por ayudarme en todo lo que pudieron a pesar de la distancia y por ayudarme a cumplir uno de mis más grandes sueños.

Finalmente agradezco a toda la gente que ha salido a las calles a exigir justicia, al pueblo chileno por haber despertado finalmente, por combatir vivazmente por el que más lo necesita, por dar la vida por la lucha. Tengo esperanza de que venceremos y que el enemigo caerá, esta vez para siempre.

A no decaer y a seguir adelante, que para eso somos buenos, el pueblo unido jamás será vencido.

Benjamín Silva

ABSTRACT

Within the field of L2, rubrics are evaluating tools that are largely used in testing procedures but have never been studied thoroughly. The influence of contexts in rubrics is another unknown territory in the area of language assessment. This investigation reports on a study of the factors that affect the quality of rubrics at the level of 10^{th} grade (*segundo medio*) in Chile. The aim of our investigation is to observe the relationship between the formation of teachers regarding rubrics and the resources given by an educational institution, and how they affect the quality of the rubric in terms of construct validity and reliability. Data was collected by interviewing the teachers in charge of 10^{th} grade from two schools of different socioeconomic context, including questions about the rubrics and about the factors we dealt with. We also analyzed two rubrics per school that were used by the interviewed teachers and evaluated them with a *meta-rubric* designed and used by the investigation team. It was found that both resources and teachers' formation had an interdependent relationship. Additionally, teachers' formation and their expert judgement seemed to affect directly the quality of the rubric, while resources in most cases affected indirectly the reliability and validity of the rubric. Time appears to be the most relevant resource as it limits the opportunities to improve the rubrics. It was also observed that teachers are critical in determining the quality of rubric design. The investigation reported could be expanded beyond the local context studied and include other stakeholders as to cover a national scope and enhance the quality of rubrics.

Keywords: rubrics, validity, reliability, teachers' formation, institutional resources.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

T1: Teacher 1

T2: Teacher 2

T3: Teacher 3

RQX: Research question x

CV: Curriculum Vitae

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

The following report describes a study aimed at describing the relationship between the resources of educational institutions, the formation, knowledge and training of teachers, and the quality of the English L2 rubrics they use.

A rubric is a scoring tool that sets the criteria for an evaluation, describing the different levels of performance in a test. Rubrics are generally composed by task descriptions, dimensions, descriptions for the levels of performance, and scales (Stevens & Levi, 2005). With rubrics, students can be aware of what is going to be evaluated and how, and teachers can evaluate them with fixed and equal criteria. Thus, rubrics help not only teachers score and give feedback to students, but also students know the contents to be evaluated in advance, with their rules and criterion, and receive feedback for future evaluations.

The study of rubrics is part of the general study of language *assessment*, "a thorough but constant appraisal, judgment and analysis of students' performance through meticulous collection of information" (Hamidi, 2010). Assessment focuses on specific points of language, and it accounts for the overall process -as opposed to evaluation, which focuses on the conclusion of a process. It is due to this that rubrics are situated in the process of language assessment, as they serve the purpose of evaluating and giving feedback for the continuation of the learning process. In this case, we are focusing on the assessment of English for Chilean students of *educación media*.

The most important properties of rubrics are validity and reliability. Validity, on the one hand, is the degree to which the theoretical and empirical evidence accounts for the right interpretation of tests and the right assignment of scores. According to Messick (1995), validity can be related to content, criterion, and construct. In summary, a rubric contributes to validity when it sets exactly what a test is measuring, when it presents an explicit and understandable criterion, and when the constructs of the abilities evaluated are empirically measured.

Reliability, on the other hand, is the extent to which the results of the evaluation are consistent and the process of scoring is replicable. When a rubric is so clearly explained and developed that it can evaluate consistently the same task with different evaluators, it is contributing to reliability. According to Brown (2004), some of the factors that can affect raters' reliability include a lack of attention to the scoring criteria, inexperience, and preconceived biases, among others. This is strictly related to the experiences and practices of the teachers or those who design and apply the rubrics, which is one of the variable factors when measuring reliability. Furthermore, all the differences that can be present among the raters could probably not be standardized if there were no rubrics.

From the perspective of stakeholders other than designers/evaluators, validity and reliability support the extent to which students can see a test/rubric as fair. This is the reason to consider a rubric as an important and useful tool when it comes to language assessment. Additionally, when teachers have a rubric, it is easier for them to give feedback to students, as its specificity can serve the purpose of explaining thoroughly the strengths and weaknesses that students showed in a language test. The more specific the rubric is, the more complete the feedback can result. Thus, rubrics contribute to both validity and reliability in the assessment process.

Furthermore, in language assessment we are also dealing with real students whose lives will be influenced by the scores they obtain during their schooling years. Therefore, the concern of how to ensure as much reliability as possible is not only theoretical and technical, but also ethical. Reliability is critic for students, because, as Muñoz (2009) phrases it, "a complex system of decisions regarding an individual's future will be based on the use and interpretation of her/his writing skills", or, we may add, any other skill that is being measured.

The importance of knowing how rubrics are designed and applied in some schools of the Chilean context is relevant as education in Chile is massively unequal. This is a consequence of the neoliberal principles of economy that were implemented in the educational system during the military dictatorship. These principles stratified the educational system, making education of good quality only accessible for a reduced group of the population with a favorable economical status. On the contrary, people without these resources just have the opportunity to access to a poor-quality educational institution, or worst, no education at all (see Section 2.5.2).

For this reason, we decided to work with two institutions of different socioeconomic contexts, in order to explore how these differences may affect in the overall result of the rubrics. In this way, due to the diverse situation in both educational institutions we chose to observe, it is essential to highlight the importance of context in the study of L2 assessment and how these differences can affect the process of assessment. This investigation will focus on describing the relation between teachers' formation, institutional resources, and the effect on the rubrics' quality in terms of construct validity and reliability. The study reported in this thesis could contribute to the understanding of how all these factors of variation and unsystematicity directly affect the validity and reliability of rubrics. Findings of the study can also guide rubric designers to a more appropriate and suitable way to measure English L2 ability in the local school contexts, beyond statistical estimation.

After making the decision of exploring the relation between rubrics and external factors, and specifying that we wanted to work at Chilean high school (*educación media*) level, we mentioned and discussed all elements that could influence the design and use of English L2 rubrics. Some factors we identified seemed related to the institutions in which the rubrics were used, and others seemed to depend directly on the teachers that used it. We also noticed that there were many other intervening factors, for example, the students' social context, parents, and other stakeholders of the institution. Finally, we grouped the factors we would work with in two categories: the resources of the institutions and the formation of the teachers in charge of *segundo medio,* the level of instruction we decided to focus on.

After the planning part of the project, all participants of the group tried to make some contacts with schools that were, ideally, different in socio-economical terms, in order to observe contrasts in terms of institutional resources and formation of teachers. Accordingly, two schools were chosen for the study, one private and one funded by a non-profit foundation. We decided to carry out an interview as the data we required was of a qualitative nature and could only be provided by teachers answering open questions. The interview included questions about the two groups of factors -resources of the institution and the formation of the teachers- and about the process of designing and applying the rubrics they used.

Before going to the selected institutions, we piloted the interview with another school in order to know how effective and understandable it was. After conducting the study interviews in the two participating institutions, the data collected was analyzed, including the information from two sources. One source was the data drawn from the answers given by teachers to the interview and the other source were the rubrics themselves. To analyze the rubrics, we also used a meta-rubric designed by us with all the information we gathered about rubrics in terms of validity and reliability (see Chapters 3, Methodology, and 4, Results, below). This instrument served as the basis to make judgments about the rubrics that we collected. Finally, both sources of data collection and their respective analyses allowed us to establish the relations we were exploring.

The study was motivated by our interest in the evaluation processes in general, and a critical observation of them. After studying the stages of evaluation, we related them to our academic processes at university and found several issues in some of the phases of evaluation, especially regarding the application of rubrics. Moreover, we were aware of the socio-economic disparity in terms of the quality of education in our country and wondered whether the different educational contexts influenced the overall quality of the rubrics. We also knew this inequity is not only present in terms of the quality of education, but also the resources and the formation of the teachers, along with the social context that students face. We found it was interesting to explore how external motives could affect the conditions of the rubrics, especially in terms of its two main properties: validity and reliability.

Chapter 2: Theoretical Framework

2.1 Introduction

The following report informs of a study conducted in order to describe the relationship between the formation of the teachers of English as a second language, the resources for the evaluating process that their educational institutions provide, and their effect in the quality of the rubrics that they use.

There is a general agreement in the literature of L2 assessment that an evaluation process must be reliable (see Section 2.4). In this sense, rubrics are tools that contribute to the reliability of an evaluation process, therefore adding validity. As it will be explained in this chapter, validity and reliability are basic components of rubrics and every valid evaluation has to be reliable.

Rubrics are not only useful to ensure reliability but also to regulate the assessment process and enhance it, which can benefit both students and teachers. According to Stevens and Levi (2005), rubrics are important because they "save time, provide timely, meaningful feedback for students, and have the potential to become an effective part of the teaching and learning process". In other words, the use of rubrics in academic contexts makes the process of evaluation and scoring more efficient for both students and teachers.

Rubrics are useful to evaluate what we deem as important in an assessment context. Inasmuch as they are a tool that allows teachers to assess precisely what they want to assess, rubrics (along with standards) play a crucial role in the development of a curriculum (Glass, 2005). Creating a curriculum, according to Glass (2005), is a four-part process which consists of identifying the standards, creating a rubric for assessment, crafting a student checklist to guide them through the unit and finally, designing lessons to achieve the standard. From this perspective, rubrics are intrinsically related with both the assessment process and the development of the design of the curriculum, functioning both as a part of these processes and as a tool to enhance them as well.

This chapter explains and discusses the theoretical framework of the study conducted on the use of English L2 rubrics in two Chilean schools. This framework includes, firstly, the definition of key concepts in the study of L2 assessment; secondly, an explanation of rubrics and their importance in L2 assessment; thirdly, a discussion of the importance of context in L2 assessment and of rubric's role in L2 assessment. In addition, the definition and importance in rubrics of validity, construct validity and reliability are also discussed. Finally, the formation of teachers and the resources given by the institutions are introduced as a definition of the context in which rubrics are used.

The chapter concludes with an introduction of the study conducted, its main objectives, and the research questions that guided it. Finally, a revision on the elements in which the quality of the rubric relies is presented, exploring further the key concepts of validity, construct validity, and reliability.

2.2 Rubrics: why are they necessary in assessment?

In this section, rubrics will be presented and discussed. Firstly, the relevance of the rubrics will be addressed, followed by a definition of what rubrics are and their role in L2 assessment. Finally, examples of rubrics will be shown in order to illustrate their role in the assessment process, as well as their components.

2.2.1 What is a rubric?

Rubrics are scoring tools that set the parameters for an evaluation (Stevens & Levi, 2005). They are divided into different components that describe in detail the levels of what is acceptable or unacceptable. Rubrics help L2 teachers by providing them with a better chance of scoring students' work and consistency. In other words, rubrics contribute to the reliability of an evaluation (see Section 2.4 in this chapter). Moreover, rubrics help students as they set the criteria of what is going to be assessed, in a language that rings true for them (Stevens & Levi, 2005). In this way, rubrics, as a scoring tool, contribute to both reliability and validity in an evaluation process.

Rubrics have a standard format. This format usually includes task descriptions, a scale, dimensions and descriptions of different levels of performance (Stevens & Levi, 2005). These components can be seen in Figure 1, taken from Stevens and Levi (2005). In this Figure, we can observe that rubrics tend to have a Title, followed by a Task

Description, Scale levels and Dimensions. Thus, there is a regular format that acts as guide for the design of rubrics.

Title

Task Description

	Scale level 1	Scale level 2	Scale level 3
Dimension 1			
Dimension 2			
Dimension 3			
Dimension 4			

Figure 1.1 Basic rubric grid format.

Figure 1: Basic rubric grid format (taken from Stevens & Levi, 2005)

Rubrics are good pedagogical tools because they come in handy when teachers need to provide timely feedback, encourage critical thinking and refine their teaching skills (Stevens & Levi, 2005). Rubrics enhance the assessment process and help teachers improve. Moreover, rubrics are quite helpful when it comes to guaranteeing respect towards equity and fairness. In fact, as stated by Stevens and Levi (2005), the main reason teachers do not use rubrics is just the fact that some of them are not aware of their existence. Despite this fact, rubrics are tools that strengthen the teaching and learning process by helping both teachers and students to have more reliable and valid evaluations.

The process of delivering feedback is boosted through the use of rubrics. According to Stevens and Levi (2005), when feedback is delivered timely it is actually more meaningful and useful for students than when it is not delivered. In this context, rubrics arise as the ideal tool for both saving time for teachers and speeding up the delivery of feedback for students, thus ensuring its effectiveness. In addition, oftentimes students do not read extended notes addressing their work and they fail to comprehend short notes. In response to this, Stevens and Levi (2005) claim that rubrics are more effective than notes as they contain as much information as possible and clearly explain what is the highest level of achievement possible. Thus, rubrics contribute greatly to feedback as they save time and contain only the information that is necessary. By doing so, rubrics helps students to improve self-assessment and self-awareness.

As a result of the use of rubrics, students are likely to develop self-awareness of their own learning process and to think critically about their improvement (Stevens and Levi, 2005). This self-awareness is what encourages them to have an active role in their education, and helps them be open-minded regarding assessment. At the same time, it also helps teachers to refine their own teaching methods, as they are able to see students' development over time. As reported by Stevens and Levi (2005), rubrics guide teachers, allowing them to gain a clear view of teaching blind spots, omissions and strengths. Thus, teachers can question their methodology and approach to teach and assess, and enhance them accordingly. Rubrics thus help teachers and students in their process of self-assessment making them agents in the learning and teaching process. These processes are developed around certain knowledge and skills that students are expected to acquire.

There is some consensus regarding the knowledge and skills that usually are assessed in an oral evaluation. In the context of this investigation, it is necessary to acknowledge them since the rubrics that will be analyzed correspond to oral tests. Figure 2, taken from Bygate (1987), describes a range of oral skills and knowledge usually taken into consideration when dealing with oral evaluations. In the Figure, he makes a basic distinction between knowledge and skill. According to this distinction, knowledge is what enables learners to talk, and skill is the active component that is involved when they are actively engaged in interaction.

Figure 5.2 A summary of oral skills (Bygate, 1987: 50)

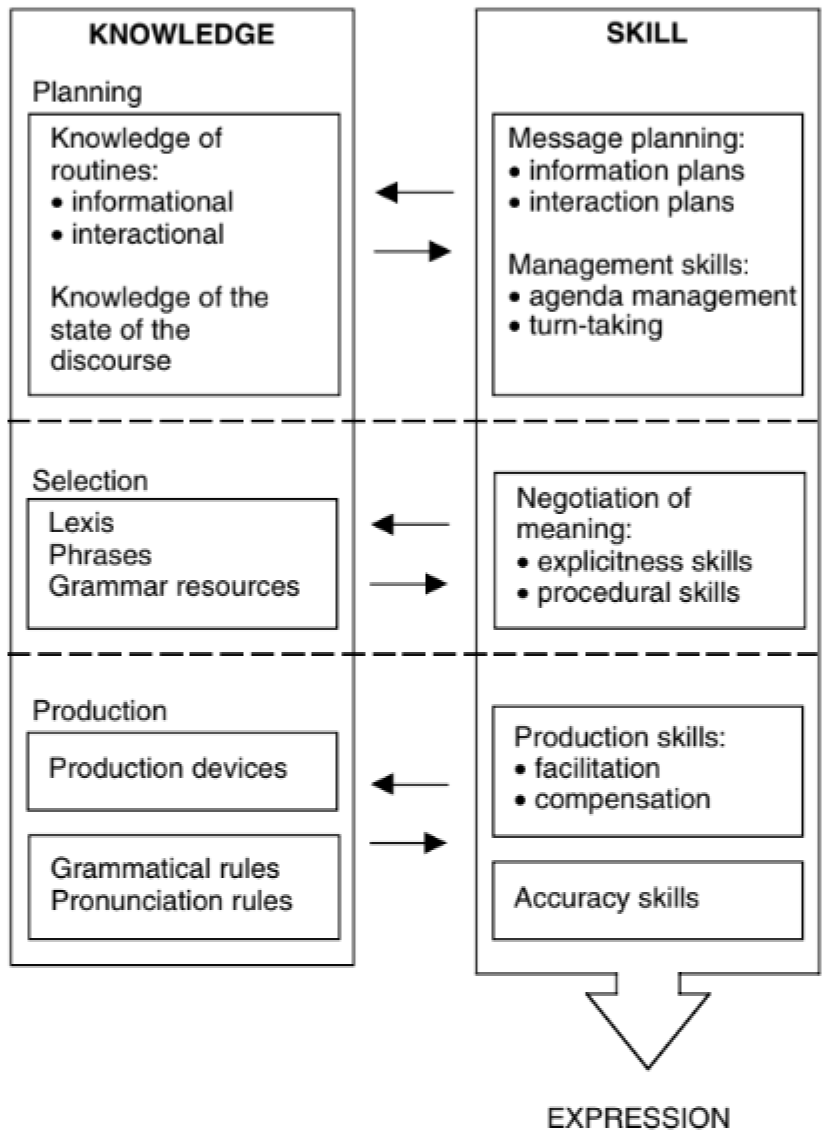

Figure 2: A summary of oral skills (taken from Bygate, 1987)

Bygate emphasizes the importance of skill practice, but recognizes that both knowledge and skill are needed when speakers speak (Luoma, 2004). When addressing rubrics that evaluate oral skills and knowledge, they usually evaluate dimensions such as planning, production, and accuracy, among others. Thus, the use of these components for the assessment process should be acknowledged in the analysis of this study. The nature of the rubric will vary according to the task and its objective.

2.2.3 The role of rubrics in L2 assessment

The practice of testing is a practice invented by humankind to make difficult decisions, and to shape educational practices and institutions (Fulcher, 2013). Moreover, testing has been used to achieve goals of control and manipulation, as well as to provide opportunities to those who would otherwise have none. Like all social constructs, it can be used for good or ill.

In this sense, it is important to review rubrics because they help in providing reliability to a process. Reliability is an instrument that tends to distribute justice among the participants of the assessment process. The distribution of justice, in turn, allows to guarantee fairness in the assessment process. Reliability and validity need to be protected since students have the right to be compared on fair grounds and to know exactly what is being assessed.

A rubric is an instrument that shows explicitly to all stakeholders what exactly an evaluation is measuring and how. According to Brown (2004), a test/rubric should clearly define the achievement that it is measuring in order to be valid in relation to its content. The ability or competence that is being measured is operationalized through the rubric with the different criteria it uses to determine the grade of ability that is at the base of the scores. All of the components that make a test/rubric valid contribute to its *face validity*, defined as "the degree to which a test looks right, and appears to measure the knowledge or abilities it claims to measure" (Brown, 2004). From this perspective, validity and reliability contribute to the extent to which students can see a test/rubric as fair. Therefore, rubrics are an important and useful tool when it comes to language assessment.

L2 assessment can impact students' lives and therefore the design and application of a rubric entails a great responsibility. As we are investigating rubrics at a local level, we are dealing with real students whose lives will be influenced by the scores they obtain during their schooling years. Reliability is critical for students, because, as Muñoz (2009) proposes regarding writing skills, "a complex system of decisions regarding an individual's future will be based on the use and interpretation of her/his writing skills", or any other skill that is being measured. By no means, therefore, should the design and application of a rubric be taken lightly.

2.2.4 Examples of rubrics

There is currently no consensus in the literature regarding the use of standardized rubrics. The discussion of this issue implies a discussion of the convenience of using standardized rubrics in local contexts over the application of a contextualized rubric. As a way of exemplifying what a contextualized rubric is, a rubric belonging to the course "Advanced English I" (Lengua Inglesa Nivel Avanzado I) is presented in Figure 3.

Rubric

	1	2	3	4
Substance	The student expresses ideas based on limited knowledge and understanding of the topic. Less than 2 sources of information are used. Ideas are underdeveloped. Cohesive devices and discourse markers are repetitive and basic.	Research on the topic was incomplete. Less than 2 sources of information are used. Ideas are expressed with some lack of clarity. Cohesive devices and discourse markers are not varied.	The student researched the topic. Arguments are supported by examples and facts from 2 or more sources, but they are not varied. The student expresses clear, well developed ideas. There is a fair use of cohesive devices and discourse markers.	The student researched the topic extensively. Arguments are supported by examples and facts from 2 or more varied sources. The student expresses clear, well developed ideas persuasively. The student uses a wide range of cohesive devices and discourse markers.
Interactive communication	The student shows little interest in the other speaker's message, gives little feedback, and asks irrelevant questions.	The student listens with partial interest in the other speaker's message; feedback is sporadic. Only general questions are asked to clarify understanding of the other speaker's point of view.	The student listens with interest in the other speaker's message, provides feedback, and asks questions to clarify understanding of the other speaker's point of view and reasoning. The speaker clarifies and acknowledges the ideas of the other speaker.	The student listens with sustained interest in the other speaker's message, provides feedback, and asks focused questions to clarify understanding of the other speaker's point of view and reasoning. The student builds on, clarifies, and gives credit to the other speaker's ideas.
Vocabulary	Very basic use of vocabulary. Use of true and false cognates. There is obvious interference with the L1.	Some of the expressions used show some degree of complexity, but this is not a constant behaviour.	The student used several sophisticated and new expressions that provided richness to his/her discourse, but at times it appeared contrived.	The student uses many sophisticated and new expressions that provided richness to their discourse making it appear natural.
Grammar	The student only produces simple grammatical structures and struggles with them. Major mistakes are made and none or only a few are self- corrected	The student produces one or two complex grammatical structures and struggles with simple structures. Some major mistakes are produced, and some are self-corrected.	Some complex grammatical structures are used but the student shows some difficulty to produce them. Some mistakes are made and the student only sometimes is able to self correct.	Complex grammatical structures are used to contribute to the points presented. Only minor mistakes are made and the student is able to self correct most of them or all of them.
Pronunciation	Intelligibility is hindered by pronunciation and intonation highly interferes with the L1.	Intelligibility is achieved, but there are several sounds from the L2 that have not been acquired. Intonation can be improved.	Intelligibility is fully achieved and only on a few occasions sounds are not produced following the patterns of the L2. Intonation is definitely closer to the L2.	
Pace and delivery	Highly hesitant. Pauses and hesitations interfere with communication. Student reads from notes more than half of the time.	Some hesitations and pauses interfere with communication. There is some dependency on the notes.	Natural hesitations and pauses filled with English sounds and expressions. No interference with communication. There is no dependency on the notes.	
Student turns	Student turns are not balanced, or the conversation is too short.	Student turns are somewhat balanced.	Student turns are balanced.	

Figure 3: Rubric sample from "Advanced English I" course

This course is part of the program English Linguistics and Literature (*Licenciatura en Lengua y Literatura Inglesas*) from University of Chile. This course was taught the first semester of the year 2019. This rubric was used as a guide by teachers to evaluate oral performances of students in a specific context and about different topics taught in classes. It also has the components that a rubric should have since it has dimensions to be evaluated, and their corresponding descriptors characterizing students' expected performances. Depending on the performance, the test is scored with a certain number of points (scale levels), that reflect the degree of knowledge/competence of the student in this area. Thus, this rubric can be considered as an example of what a traditional rubric is.

The example in Figure 4 below corresponds to a standard rubric created by Cambridge English: Advanced (Level C1) Speaking to evaluate the speaking performance of students that belong to the C1 level. Here we can see five dimensions that are being evaluated and five different levels. Each dimension has its respective descriptor that explains in detail what is expected in terms of performance. This rubric is public and free-to-use, hence, teachers can easily search for this rubric so that they can implement it in their courses. Candidate speaking performances are assessed using scales which are linked to the Common European Framework of Reference. The assessor gives 0–5 marks for each of the criterion (English, tests & Advanced, 2019).

C1	Grammatical Resource	Lexical Resource	Discourse Management	Pronunciation	Interactive Communication
5	• Maintains control of a wide range of grammatical forms.	• Uses a wide range of appropriate vocabulary to give and exchange views on familiar and unfamiliar topics.	• Produces extended stretches of language with ease and with very little hesitation. • Contributions are relevant, coherent and varied. • Uses a wide range of cohesive devices and discourse markers.	• Is intelligible. • Phonological features are used effectively to convey and enhance meaning.	• Interacts with ease, linking contributions to those of other speakers. • Widens the scope of the interaction and negotiates towards an outcome.
4	*Performance shares features of Bands 3 and 5.*				
3	• Shows a good degree of control of a range of simple and some complex grammatical forms.	• Uses a range of appropriate vocabulary to give and exchange views on familiar and unfamiliar topics.	• Produces extended stretches of language with very little hesitation. • Contributions are relevant and there is a clear organisation of ideas. • Uses a range of cohesive devices and discourse markers.	• Is intelligible. • Intonation is appropriate. • Sentence and word stress is accurately placed. • Individual sounds are articulated clearly.	• Initiates and responds appropriately, linking contributions to those of other speakers. • Maintains and develops the interaction and negotiates towards an outcome.
2	*Performance shares features of Bands 1 and 3.*				
1	• Shows a good degree of control of simple grammatical forms, and attempts some complex grammatical forms.	• Uses appropriate vocabulary to give and exchange views, but only when talking about familiar topics.	• Produces extended stretches of language despite some hesitation. • Contributions are relevant and there is very little repetition. • Uses a range of cohesive devices.	• Is intelligible. • Intonation is generally appropriate. • Sentence and word stress is generally accurately placed. • Individual sounds are generally articulated clearly.	• Initiates and responds appropriately. • Maintains and develops the interaction and negotiates towards an outcome with very little support.
0	*Performance below Band 1.*				

Figure 4: Sample rubric of C1 level from Cambridge English: Advanced (level C1) Speaking.

2.2.5 Quality of a rubric: how is it affected by validity and reliability?

One main assumption of this study is that the quality of a rubric relies on its construct validity and its reliability. *Construct validity* is the proof that the skills and knowledge that should be evaluated are actually the ones being evaluated (Messick, 1995). On the other hand, reliability implies that the descriptors are clear on what has to be assessed, so that nothing depends on teachers' individual judgment. Reliability also implies that the levels of attainment are well developed, making it easy for students to grasp the difference between the levels.

2.3 Validity: definition and how to ensure it

According to Messick (1989, in Bachman, 1990), validity is "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores." (p. 34). Messick considers validity as a measurement to get empirical facts to endure suitability and adequacy to test score competences.

Unsatisfied with the definition mentioned above, Bachman provides the definition of validity by the American Psychiatric Association (APA) from 1985: "Validity always refers to the degree to which that evidence supports the inferences that are made from scores." This definition assumes therefore that the empirical evidence and theoretical rationales support the scoring nature of validity.

This definition complements Messick's definition in the sense that validity is about evidence that supports results obtained from test scoring. Similarly, the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999) presents definition of validity focused on evidence: "the degree to which evidence and theory support the interpretations of test scores entailed by the proposed uses of a test." (p. 9). In this case, the aspect of theory is added to the interpretation of scoring. This aspect was not considered directly in APA's definition of validity.

In line with these definitions, Bachman (1990) concludes that validity can be demonstrated only through the collection and interpretation of all significant sources and types of information. Messick (1995) explains that the traditional view of validity consists of three types of this attribute: *content validity*, *criterion validity*, and *construct validity*. Content validity, according to Brown (2004), is identifiable when one can "clearly define the achievement that you are measuring" (p. 22). Criterion validity, in turn, has to do with the criterion of the test being reached (Brown, 2004). Finally, construct validity deals with the extent to which a construct is empirically measured (Brown, 2004). This last type of validity will be explained further in the next section of this study (Section 2.3.1).

These three sources of test validity are not the only types of validity we can find. In addition to the types already mentioned, there is also *consequential validity*, which deals with all the consequences of applying a test (Brown, 2004), and *face validity*, which refers to the extent to which the test appears to "look right" and measures everything that it claims to do (Mousavi, 2002, in Brown, 2004).

Messick (1995) argues that the traditional view of validity is incomplete, since it does not take into account the value implications of score meaning as a basis for action and the social consequences of score use. He suggests, instead, a more complete view, which he calls "unified validity". Unified validity, according to Messick (1995), integrates content, criteria, and consequences into a construct framework. This concept can be applied or scientific, to test rational hypotheses about score meaning. In Messick (1989b, in Messick, 1995), validity is described as:

> "An overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions of the basis of test scores or other models of assessment." (p. 741)

In other words, he sees validity as the extent to which we can find empirical evidence to support how appropriate interpretations and actions can be based on test scores. Messick (1995) clarifies that validity is a meaning of the test scores rather than a property of the assessment, since what needs to be valid is the meaning or interpretation of test scores. Furthermore, validity is an evolving property due to the fact that meaning may vary across different people, groups, settings, or contexts. Therefore, the validation of a test is a continuing process.

According to Messick, validity is not only a measurement principle, but also a social value that has meaning and force outside measurement whenever judgments and decisions are made (1995). Thus, being a social value, validity has also a political and scientific role, in that evidence and proof is required for assessments to be relevant, combining scientific inquiry and rational arguments to both justify or nullify score interpretation and use.

Messick (1995) also defines validity as “nothing less than an evaluative summary of both the evidence for and the actual consequences of score interpretation and use” (p. 742). All in all, construct validity integrates any source of evidence that can be used in the interpretation or meaning of test scores. Messick argues that validity evidence is the basis of test validity, since in order to know whether a test is valid or not, one needs to have evidence that serves as anticipation of consequences first (1995). As stated above, evidence is in the nature of construct validity.

2.3.1 Construct Validity

Construct validity is defined in Souza et al. (2017) as the degree to which a group of variables represents the construct to be measured in a study. In order to establish construct validity, the academic has to construct a hypothesis and test them to support the instrument validity. Regarding construct validity, Bachman (1990) discusses the analysis of the processes underlying test performance. On the one hand, he clarifies that quantitative analyses of patterns of test performance and the relationships between that and the treatments of test performance are very limited in their capacity for creating new hypotheses. He also mentions that a more specific limitation for construct validity has to do with correlational and experimental approaches to construct validation. This is because such approaches examine only the products of test taking (scores) and “provides no means for investigating the processes of test taking themselves” (Bachman, 1990).

As described in Section 2.3 above, validation implies empirical evaluation of the meaning and consequences of measurement (Messick, 1995). The term “empirical evaluation” means that the process of validation requires both scientific and rhetoric evidence, (i.e. studies or investigations and rational arguments) in order to be deemed as valid. To have this evidence, six aspects of construct validity are required. According to Messick (1996), these six aspects are crucial to emphasize the central issues regarding unified validity. They “function as general validity criteria or standards for all educational and psychological measurements” (p. 7), and they should not be presented as separate and different kind of validity, such as the ones mentioned in section 2.3. Instead, they should be viewed as "interdependent and complementary forms of validity evidence" (p. 7).

First, there is the *content* aspect, which consists of the evidence of content relevance, representativeness, and technical quality. Second, the *substantive* aspect, which refers to the theoretical rationales for all consistencies observed in test responses, including process models of task performance and empirical evidence that proves that the theoretical processes are in fact engaged by respondents in the task. Third, the *structural* aspect, which has to do with the fidelity of the scoring structure to the structure of the construct domain. The *generalizability* aspect deals with the extent to which score properties and interpretations can be generalized across groups, settings, and tasks. The *external* aspect refers to the extent to which the scores' relationship with other measures (convergent pattern) and non-assessment behaviors (discriminant pattern) reflect the expected results that are implicit in the theory of the construct that is being assessed. Finally, the *consequential* aspect has to do with value implications of score interpretation as a basis for action. It also refers to the actual consequences of test use regarding sources of invalidity, such as bias, fairness, and distributive justice (Messick, 1995).

According to Messick (1995), the essence of unified validity is that appropriateness, meaningfulness, and usefulness cannot be separated due to the empirically based score interpretation. If validity evidence is obtained, value implications and social consequences are to be brought into the framework of unified validity. This framework distinguishes two facets that are interconnected: one source of justification of either evidence supportive of score meaning or consequences contributing to score valuation, and a second one, of the function or outcome of the testing, either interpretation or applied use. Thus, a four-fold classification of unified validity can be obtained, since these facets overlap each other.

	Test interpretation	Test use
Evidential basis	Construct validity (CV)	CV + relevance/utility (R/U)
Consequential basis	CV + value implications (VI)	CV + R/U + VI + social consequences

Figure 5: Classification of unified validity (taken from Messick 1995).

Firstly, in Messick (1995)'s classification, we find evidential basis for test interpretation, where evidence acts out as a means of trustworthiness of score meaning, which in turn transforms evidential basis into construct validity. As for test use, its evidential basis is also construct validity, in that the evidence of score meaning can either include or become enhanced by evidence of the relevance of scores to the applied purpose and for the utility of the scores in the applied setting.

The consequential basis of test interpretation integrates not only construct validity, but also value implications of score meaning, which are socially relevant for applied practice and social policies. On the other hand, the consequential basis of test use has to do with assessing the value of both the potential and actual social consequences of the testing by taking into account benefits and risks of the test use that was proposed and comparing them with the pros and cons of its counter-proposals. By doing this, one should gain input (i.e., new ideas) that could be beneficial for the actual proposed test use. What is important to have in mind is that it does not matter whether the consequences are positive or negative, but rather how they came to be and what exactly determined them. Most importantly, these social consequences should not be what makes the test use invalid if they are adverse, they should not be attributable to any source of invalidity, which is why the last cell in Figure 5 integrates construct validity, relevance and utility, value implications and social consequences (Messick, 1995).

In sum, rubrics' construct validity is composed by evidence that may come from different sources, such as the ones already stated. Above all, evidence in validation, whether scientifical or rhetoric, contributes to the construct validity of rubrics more than to content and criterion validity. This is so because, in order to have construct validity, one needs empirical evidence of what is being measured, which is not something that is required in the other types of validity.

2.4 Reliability

Many definitions of reliability can be found in the literature of assessment studies. This section presents the ones that seem to be most relevant. Additionally, some examples of reliable measurements and their corresponding explanations will aid to illustrate the functions of reliability.

In everyday life, to be reliable is quite similar, if not a synonym, to being "dependable" or "trustworthy". Nonetheless, when referring to evaluation, to be reliable requires a much more precise definition, as a concept such as "dependable" can easily be confused with the idea of validity (Trochim, 2006). On that account, many definitions consider reliability to be equal to "repeatability" or "consistency" (Jonsson & Svingby, 2007). According to Trochim, for instance, "a measure is considered reliable if it would give us the same result over and over again", thus ensuring consistency in assessment.

In general terms, the property of reliability could be illustrated with the functioning of a simple object of everyday life: a clock. This timepiece has the obvious characteristic of informing hours, minutes, and seconds, and it is of course expected to be consistent in the way it measures time. In other words, even if it is not synchronized with other clocks, there should always be the same amount of time between hours, and the instrument should repeat the same patterns of time measurement every day. If this were to vary, the clock would be inconsistent in its measurements, and therefore unreliable.

For assessment to be reliable, scores are expected to be equally consistent as a clock or any measurement tool. This can be observed, for example, when a student takes two times and the obtained results are either similar or the same (Brown, 2004). Such similar results should be obtained independently from external variables such as when the student performed the test, when it was evaluated, and who evaluated the student's test (Moskal & Leydens, 2000). On the contrary, if the score varies because of factors that are unrelated to the real purpose of the assigned task, the test is considered unreliable.

The idea of reliability belongs more generally to the concept of "replicability" which, according to Muñoz (2009), is another way to define reliability. Under this concept, an examination is reliable when test scores have the ability to be replicable from one occasion to another. Muñoz poses that this definition has four key implications: first, that reliability is a statistical property of test scores; second, that the scores show consistency at different assessment occasion; third, that there is a way to measure this consistency; and fourth, that it is a necessary condition for any kind of assessment, although it is not sufficient by itself, just like validity (Muñoz, 2009). As can be seen, reliability and consistency work hand in hand and are basically indivisible. In simple terms, "the more consistent the scores are over different raters and occasions, the more reliable the assessment is thought to be" (Moskal & Leydens, 2000).

It can thus be stated that rubrics contribute to the reliability of an assessment process as they are instrumental in the evaluation of the same task or work by different evaluators and/or on different occasions. Rubrics have the criteria that should establish descriptors and dimensions known by the evaluators. A series of practices can be implemented, such as calibrations, and instruments that help them to evaluate performances equivalently.

Arguably, the nature of the evaluation will vary depending on the type of reliability. This differentiation has been already stated when claiming that evaluations can be reliable both when there is agreement in the assessment of the same task graded by two or more raters, and when there is consistency in the assessment of the same task in different occasions by the same rater(s).

2.4.1 Types of reliability

As reported by Trochim (2006), there are four classes of reliability and ways in which we can estimate reliability for each of them. These are inter-rater or inter-observer reliability, intra-rater reliability, test-retest reliability, parallel forms reliability, and internal consistency reliability.

Regular educational contexts usually present two types of reliability: inter-observer and intra-rater reliability. Both of them will be presented due to their current appearance within the process of evaluation. Therefore, in this investigation we will be introducing only these two types of reliability.

Inter-rater reliability can be observed in assessment occasions where tests can be scored by more than one rater. Bearing in mind that variations in judgment can occur across raters (Jonsson & Svingby, 2007), the aim of inter-rater reliability is to ensure people are not inconsistent amongst their scores while measuring a sample. To estimate inter-rater reliability, it is suggested to calculate the agreement between the raters or calculate the correlation between them. As for intra-rater reliability, the raters' judgment can vary "in the inconsistency of one single rater" (Jonsson & Svingby, 2007).

2.4.2 Theoretical foundations of reliability: true value and error

Trochim (2006) poses the basic concepts that define the theoretical basis of reliability. First, the *true score theory* is at the basis of the concept of reliability. Second, two types of measurement error exist: *random* and *systematic*.

According to the true score theory, all measurements are composed by the true value, or true ability that is being measured, plus some random error value. This is summarized in the following equation:

$X = T + ex.$

The model poses that most measurement has an error component. This error can be due to personal factors, such as anxiety, or to environmental factors, such as noise. Because assessment occasions are mostly subject to variation, the true score is a more realistic way to measure reliability (Trochim, 2006)

As stated by Trochim, (2006) the error components in an assessment can be *random* or *systematic*. On the one hand, random error is caused by factors that randomly affect the measurement, for instance, someone's mood or context may inflate or deflate their performance. On the other hand, systematic error affects the entire measurement of the variable. An illustration of this would be the noise present in the environment in which a test is being taken. In this case, all students could be distracted from doing the task. In simple terms, random error does not have a consistent effect across the entire sample as systematic errors do.

In order to reduce measurement errors, either random or systematic, in the Web Center for Social Research Methods (Trochim, 2006) it is claimed that some strategies can be put in practice. First, you can pilot your instruments, hence, you can get feedback about the difficulty of the measure and about how the environment affected the development of the entire measurement. Second, if data is being collected by people, you should train the interviewers or observers to make sure they are not introducing error. Third, you should always double-check the data you collected. Fourth, you can use statistical procedures to reduce measurement error. Finally, you can use multiple measures of the same construct to be able to triangulate across the different measures.

The possibility of either systematic or random error poses the question of how to actually measure reliability by using statistical procedures that aim to reduce measurement error. Trochim (2006), for instance, claims that reliability cannot be calculated. Instead, it must be estimated. Along these lines, Muñoz (2009) explains reliability as "the extent to which differences in ability are not obscured by random measurement errors". Such a definition implies that, if there were no measurement error, a test should ideally correlate or agree with itself or another test. Ultimately, this means more reliability for the assessment procedure (Muñoz, 2009). Since error is inevitable in human judgments, the fact that language assessment is always subject to some degree of variation must be considered when statistically estimating reliability. Accordingly, Bachman (1990) argues that "the demonstration of reliability consists of estimating the amount of variation in language test scores that is due to measurement error".

In conclusion, reliability in assessment means that scores are expected to be equally consistent, whether there are two or more evaluators or a single one. This distinction will be presumably reflected in the design and the quality of the rubric. In both cases, reliability can be estimated with the aid of statistical procedures. Furthermore, variation inevitably implies either *random* or *systematic* measurement errors. This must be born in mind when estimating reliability in rubrics, for which the model of the *true score theory* is handy. Reliability is thus key in the assessment process, and more specifically, in the design of rubrics. This, of course, contributes to validity, as both of them are crucial factors that affect the quality of the rubric.

2.5 Factors that affect the rubric

The presence of construct validity in terms of ensuring that the rubric will evaluate the task stated by the teacher, and reliability in that every student will be evaluated in a consistent way and by the same terms will ensure the quality of a rubric. Based on this definition, it is important to acknowledge the variables that affect the quality of a rubric.

In spite of the quality of rubrics, we cannot forget that every instance of evaluation has to be judged within context. Brown (2004) considers that the issue of reliability of a test in context. He claims that "the issue of reliability of a test may best be addressed by considering a number of factors that may contribute to the unreliability of a test" (p. 20). Within reliability, the author points out two aspects of assessment that are affected in the context: student-related reliability and rater reliability.

On the one hand, Brown (2004) discusses that the most common learner-related issue is a consequence of personal problems such as temporary illness, fatigue, and anxiety, among other physical and psychological issues that may make the "observed score deviate from one's true score" (p. 21). On the other hand, human raters' mistakes, subjectivity, and bias may also interfere with the scoring process. Inter-rater reliability, on the other hand, is described as an occasion when "two or more scores yield inconsistent scores of the same test, possibility for lack of attention to scoring criteria, inexperience, inattention, or even preconceived biases" (p. 21). Therefore, reliability in rubrics is limited by people's capacities and the ability to overcome the situations mentioned above. In order for rubrics to be valid, awareness of these factors is necessary.

According to Muñoz (2009) it is known that human unsystematicity is a source of random error in the estimation of reliability: due to contextual and personal factors, actors perform differently in different occasions. As for the latter, factors could include tiredness, anxiety, and the actor's personal background, among others. Environmental causes, meanwhile, refer to factors such as room temperature, noise, available infrastructure, and institutional resources. Regarding raters, their decisions are shaped by personal beliefs, preferences and biases, and their differences in judgment answer to testing conditions or individual features. Therefore, rubrics are hard to apply into contexts that are dynamic in nature. Therefore, teachers have to take into account how to adapt their rubrics into their context.

It has been already stated that context has an important role when it comes to ensuring reliability, and therefore quality, in a rubric. Among the circumstances mentioned above, there are other contexts that also affect the quality of a rubric. One of them can be referred to as *human context.* This specific context includes students, parents, curricula supervisors, and teachers. The second context can be understood as *material context.* This context includes "Academic load" and "Infrastructure and equipment". When we speak about academic load, we refer to factors such as how many the hours the teachers have to work on rubrics, the quantity of students per teacher, and finally the quantity of different processes of evaluation that the teachers need to work on. On the other hand, when we refer to infrastructure and equipment, such as physical materials that the teachers use to design and apply their rubrics, like pedagogical materials (textbooks, glossaries, etc.), physical space, and technological equipment such as a computers, tablets, and Internet.

2.5.1 Formation of teachers and human context

There are different groups of people who are interested in the process of designing, applying, and using rubrics. First and foremost, both teachers and students are the ones who interact with these tools directly. At the same time, there are also stakeholders including school administrators, who play a part in this process. Indeed, there are times where it is not the teacher who designs his or her own rubric, but a stakeholder, usually from the school, who does the initial designing. Although ultimately it is the teacher who eventually modifies and applies a rubric, there are more people involved in this process. We will call this human context, and we will consider as such all of the people who get to interact with the rubric, whether that may be in terms of design, application, or usage.

In order to better understand what the human context is, we need to acknowledge Davidson and Lynch (2002: 76-78, in Fulcher, 2013)'s usage of the term '*mandate*'. This term alludes to "where the test purpose comes from" (p. 1). For the purposes of this study, we will also extend this view to rubrics.

Mandates can be either internal or external. Internal mandates usually refer to those tests that are established by the teachers themselves or by school administrators who are aware of the students' context and their environment. This awareness of the context allows this kind of mandate to be understood as '*ecologically sensitive*' (Fulcher, 2013), as it serves the particular needs of both teachers and learners in context.

On the other hand, external mandates are all of those tests that come from outside the context of the school in question. Under these circumstances, teachers from the establishment do not have any saying in the making of these tests, since they are imposed on them. Such is the case of tests that can be generalized, like standardized tests.

Since there are more people apart from teachers who are responsible for the making of rubrics, it is crucial that teachers, who are the ones that apply rubrics, know the way in which rubrics work. This knowledge can help them to ensure the assessments' validity and reliability. However, there is no way of truly ensuring that every teacher of English as an L2 has the same level of formation regarding rubrics and evaluation. What is more, according to figures provided by SIDOC (*Sistema de Gestión Documental*), almost half of the people who currently teach English are not teachers (Vivanco, 2012). The fact that these people are not teachers suggests that their formation is not aimed at pedagogy. Being this the case, it can be assumed that the knowledge educators may possess about evaluation and rubrics, in particular, is limited.

By not having the proper formation or specialization as teachers, it can be expected that these educators will have scarce knowledge about the process of designing and applying rubrics (as shown in Harris & Eggen, 1993). As we have already noted in sections 2.3 and 2.4, these two aspects of a rubric are key for every evaluation. Nevertheless, through their experience teaching they can develop an expert judgment, which is defined as "una opinión informada de personas con trayectoria en el tema, que son reconocidas por otros como expertos cualificados en éste, y que pueden dar información, evidencia, juicios y valoraciones" (Escobar-Pérez & Cuervo-Martínez, 2008). In turn, both their formation and expert judgment can be expected to affect the overall quality of the rubric in terms of validity and reliability.

2.5.2 Resources given by the institutions and material context

As we have already mentioned, there is a human context, which alludes to the teachers, students, and stakeholders of the school, and a non-human context, which deals with the institutions' resources when considering the quality of rubrics. However, to understand the non-human context, we will need to contextualize the model of Chilean education and the reforms it has had in the past decades, as well as the laws that have been created and implemented in relation to the educational system. This contextualization is necessary due to the fact that we will be taking into account resources from Chilean schools specifically, and the model that reigns in Chile is characterized by extreme differences in the resources that institutions have.

To begin with, according to Cornejo (2006), in the beginning of the 1980's, and in the context of the military dictatorship, Chile went through a radical educational experiment. This experiment consisted in the application of basic neoliberal principles to the educational policies of the time (p. 118). This model had two big structural pillars, one regarding the radical transformation of the way in which the schools were financed (through "*vouchers*" or grants), and the other one regarding the Model of Management of the Educational System. The latter process was also known as *municipalization* (Cornejo, 2006).

According to OCDE (2004: 177, in Cornejo, 2006), this financial model consisted in the State giving grants to private schools in the same way they give grants to municipal establishments. The idea behind this was to introduce mechanisms of competition similar to those of the market to schools. The establishments would compete to find students so they would be given the grant from the State. In doing so, the academic results of the schools would improve (p. 119).

There are three types of schools in Chile. As we have already mentioned, there are public schools that are funded by the different municipalities, private schools that are given grants by the State, and finally private schools that are not given grants by the State. Moreover, there are three levels of management and administration. These are the central level, which corresponds to the Ministry of Education; the intermediate levels, which have the role of connecting the schools with the central level; and the management teams of each educational establishment. These three levels are what we call the model of management of the educational system (Cornejo, 2006). The purpose of this scheme was the decentralization of the current model, by allowing the municipalization of schools, where schools depended financially of stakeholders that could be provided by the municipality or by their own private means (p. 119).

One of the main difficulties that this model presented, according to Cornejo (2006), was that the so-called stakeholders did not provide any technical pedagogical equipment that allowed the making of decisions of schools. Instead, only decisions regarding the monetary gain of the establishment were most likely to be taken. What is more, from 345 districts, only one third counted with stable professional support equipment, and from that third of districts, only 32 had more developed and complex equipment, which clearly translated to those municipalities with better financial context (Montt, 2004, in Cornejo, 2006). This is an important issue to consider, since having these kinds of differences between establishments from different districts supposes a structural source of inequality in education, due to the lack of resources some schools have.

As a consequence of this context, Cornejo (2006) argues that both the school and the teachers lack of support and quality technical supervision. Furthermore, it is true that students still have access to education, but the quality of the education they receive will equate to the financial status of the school. In Cornejo's words, "the quality seems to be reserved to those who can pay for it" (2006: 125).

Based on the work done by Alberto Moreno and Rodrigo Gamboa (2014), we could find that the inequality on the educational system in Chile comes from the 1973 Dictatorship and from several laws from the Constitution of 1980. It is a fact that "Chile has a segregated educational system translated into schools for poor people, middle class and rich people, situation that has been reported for international organizations" (p.52)

Among all the elements that affected directly and indirectly the educational system in Chile, there are three laws created and based on the Constitution of 1980 that helped maintaining the market ideas towards education. These laws led us to a better understanding that education has not the same conditions for everyone, thus, it is unequal.

Firstly, there is the law of academic freedom, which gives liberties to any person to create a school based on a few characteristics given.

> "La libertad de enseñanza incluye el derecho de abrir, organizar y mantener establecimientos educacionales. La libertad de enseñanza no tiene otras limitaciones que las impuestas por la moral, las buenas costumbres, el orden público y la seguridad nacional. La enseñanza reconocida oficialmente no podrá orientarse a propagar tendencia político partidista alguna. Los padres tendrán el derecho de escoger el establecimiento de enseñanza para sus hijos" (CHILE, 2002, p. 13 in Moreno and Gamboa, 2014 p.55).

This situation led to what we know as *market education*. As explained before this law created several institutions funded by private supporters, partially funded by the State and the other ones had a different future. An important milestone occurred when education went through a process of decentralization and education became municipalized. After this, it was stated that "the Ministry of Education will have the sole administrative responsibility for the curricular issues and compliance of student attendance" (p.55)

Secondly, a second law regarding education was developed. This referred to the payment of teachers and modified characteristics of the regulation of work policies. In this term that "the law promulgated DFL Nº 1.3063of 1980 presents in its article 4, in addition to announcing the logic of decentralization, that the Labor and Social Security Code will be governed by what is proposed by the private sector, leaving the operation of educational centers to the criteria of supporters and municipalities" (p.55). This law has direct implications on the integral security of teachers who work on the State educational system because now it depends on the municipalities whether or not they are being paid and how. This situation serves as further proof that education lacks of equality based on the idea that there are certain districts that have more money than others, therefore there will be more resources in certain schools.

Finally, the proof that education has unequal conditions is the one concerning the basis for the curriculum within schools. This law says that "it is authorized, through the law promulgated Nº 4002 of 1980, the reduction of working hours and the elimination of any subject due to the inability to cover due to economic causes" (p.56). This change is important because of the differences between public and private education that began to show. Plans and programs were transformed into the "the mandatory minimum for institutions with more resources" (p.56), which remains the top education for private schools and demonstrates the marginalized situation that is still remaining on public education in Chile.

As a conclusion, a gap between schools of fewer resources and schools which have more resources can be found. On the one hand, a school with more resources will be able to hire more teachers, which will reduce academic load. Moreover, private schools may be able to acquire teachers who are better prepared in terms of rubrics and evaluation. In contrast, a school with fewer resources will only be able to hire a certain lower quantity of teachers, who will in turn be in charge of more classes and more students, which limits their time to work in rubrics. Finally, an establishment with more resources will also be able to provide the teachers with more planification hours than an establishment with fewer resources. Because of this, the teachers will have less time to design and develop rubrics.

Even if there is proof in the literature (Cornejo, 2006; Davidson and Lynch, 2002, in Fulcher, 2013; Moreno and Gamboa 2014; Vivanco, 2012) that the aforementioned factors, namely the human and non-human context, do affect the education that is being imparted to young learners in terms of tests, we found that there was no problematization of the issue in terms of rubrics. In terms of resources such as the infrastructure and equipment of the establishments, no mentions of these actually affecting the rubrics of schools were found.

Nevertheless, we do know that the current educational system in Chile presents a great degree of inequality, where the quality of the education imparted is being subjected to the financial stability, or lack thereof, of the establishment in terms of the quality of the teachers and the infrastructure at disposition. Having this in mind, we can assume that these factors do affect rubrics as well as the tests of every single establishment in Chile.

2.8 A study of rubrics in context

Several factors can affect the overall quality of a rubric, especially in terms of validity and reliability. Rubrics do not function in isolation; these instruments are constantly influencing and being influenced by external elements that can depend to different extents on those who apply the rubrics. Human and non-human contexts can provide infinite possibilities of circumstances in which rubrics are designed and applied. These factors, however, are not problematized in the existing literature about rubrics and evaluations in general, which leaves us with an important knowledge gap to fill in.

This gap is even broader when we place ourselves in the Chilean context. On the one hand, education is not state-owned but depends on the different institutions that provide it, which can lead to an unequal distribution of resources, as well as a profound differentiation of methods and programs. On the other hand, teachers and stakeholders are not trained and qualified to the same extent and in the same way, so there is no warranty that they can design and apply rubrics with enough and equal capacities.

All in all, there are various scenarios in which rubrics can be developed, and a significant number of factors that can affect their quality. The present study aims at accounting for the relation between some of these factors and their effect in a set of rubrics. In this section, we are presenting the relevant information about the study. Firstly, the objective of the study, then the research questions, and finally the procedures of the study.

2.8.1 Objective of the study

The study reported aims at describing the relation between the resources of an educational institution and the formation and knowledge of its teachers, and the quality of the rubrics used by them. This is a case study in which two different schools are observed as cases, analyzing the rubrics used for English classes in *segundo medio* (equivalent to 10th grade of High School in the United States).

2.8.2 Research questions

The research questions that guided the study were established as follows:

1. Which is the formation and knowledge that teachers of English as a second language have at the moment of developing and evaluating rubrics?

2. What are the institutional resources that the teachers can use for the design and application of the rubrics?
3. Which is the way in which teachers design and apply the evaluating rubrics to their students?
4. What is the relation that exists between the quality of the rubric and the interaction between the formation and knowledge of the teachers, and the institutional resources in relation to construct validity and reliability?

The study was conducted by visiting the two schools selected in order to carry out an interview with the English teachers of *segundo medio*. This interview included questions about the resources of the institution, the formation and knowledge on rubrics and evaluation of the teachers, and the way in which they designed and applied the rubrics. The information given by the teachers in the interview provided the data to answer questions 1 to 3 of the study. On the other hand, the teachers provided us with the rubrics they used in their classes, the tests that corresponded to each rubric, and the program of the year. These documents were revised in order to analyze the rubrics' features and assess their alignment with the tasks and the programs. The characteristics and quality of the rubrics were extracted from that analysis, and conclusions from the interaction between such characteristics and the resources of the institutions and the capacities of the teachers were drawn in order to answer the fourth question of the study.

In the next chapter of this study, the methodology used to develop this investigation will be presented. Firstly, we will go through participants. Secondly, we will address data collection. Finally, we will present data analysis procedures.

Chapter 3: Methodology

The purpose of this study is to explore the interaction between the formation of teachers the resources they receive from institutions and the effects of this interaction in terms of the impact it has on the quality of the rubrics they use.

Thus, the objective of the study reported in this study is to describe the formation and knowledge of teachers for the evaluation and development of English L2 rubrics, as well as the institutional resources that they receive in their particular contexts.

Another objective is to describe the process of design and application of rubrics used by teachers on characterizing the relationship between the quality of the rubric and the interaction of both formation and knowledge of teachers with the resources that educational institutions provide to them. The quality of the rubric has been established in terms of their contribution to the construct validity and reliability of the assessment processes where rubrics are included. (see Section 2.2 in Chapter 2)

Some of the procedures involved in the development of this investigation were qualitative. According to Phakiti (2015), the basic foundation of a qualitative study is to investigate in depth every aspect related to a behavior or social practice. A qualitative approach is thus adequate when studying the development of the quality of the rubrics made by teachers in particular educational institutions. Bachman (1990) mentions that researchers have begun to use qualitative research procedures in order to investigate test taking. This is the case of the current investigation, which examines oral tests.

In this chapter, we will present the information about the participants that were part of the study, as well as the data collection procedures used to carry out the study. We will explain the procedures applied for the analysis of the data collected in order to answer the research questions of the study (see Section 3.3.1 of this chapter). Data collection tools will also be described, together with an introduction of the interview and of the rubrics provided by the interviewed teachers and a meta-rubric created by the research team in order to analyze, describe, and evaluate the rubrics given.

3.2 Data Collection

In this section, we will start by presenting and describing the data collected, including interviews, rubrics, and course programs, as well as the way they were collected. Additionally, the tools used to analyze the data will be described and explained. The meta-rubric can be found here, which we used in order to evaluate the different aspects rubrics, according to us, should possess. It must be mentioned that the meta-rubric is not an instrument to collect data, but to analyze it. The interview was conducted to get information about teachers' formation regarding the design and application of rubrics and the institutions' resources through the content found in the answers (see analysis in Discussion of Results in Chapter 5).

3.2.1 Participants

Participants of the study were three high school (*educación media*) teachers from two different educational institutions. One teacher taught in Escuela Arturo Toro Amor, which is located in Independencia district and the other two teachers worked together in Scuola Italiana, located in Las Condes district. Each participant was interviewed regarding the design and application of the rubric, their formation on evaluation and rubrics, and the resources that each institution provides them for the use and application of rubrics (see interview details in Appendix in Chapter 7).

Participants were selected in agreement with the main objective of the study. The objective of the study was to establish a relationship between both institutional resources that the teachers have and their formation on evaluation of the design and the application of rubrics. Based on the fact that ultimately teachers are the ones involved directly in the process of design and application on rubrics, we will be focusing on them as the main factor that affects the quality of a rubric. In this context, we chose to work with these two institutions due to their differences regarding their funding, location, and the number of teachers they have. Furthermore, the differences between these institutions were expected to be associated with the level of inequality in terms of resources since one of the institutions has more resources than the other (see Section 4.2).

The selection of participants/ institutions was expected to provide an opportunity to observe similarities and differences regarding the relationship of the three variables we were investigating, namely the formation and previous knowledge of the teachers, the institutional resources that they are provided with by the school, and the relationship between these two in terms of the quality of the rubric. It was also expected that the cases would allow us to find out if the construct validity and reliability of the rubrics used by participant teachers might be related to the differences between these institutions.

3.2.2 Data collection tools

A semi-structured interview was designed with the aim of collecting the necessary data from teachers of different institutions regarding their own experience; the resources they have available to learn about rubrics and develop rubrics; the way in which they design and apply rubrics; and the feedback they finally give to students.

Semi-structured interviews are characterized for having "a set of pre-prepared guiding questions and prompts, the format is open-ended and the interviewee is encouraged to elaborate on the issues raised in an exploratory manner" (Dörnyei, 2007). According to Dörnyei (2007), an interview is a natural and acceptable way of collecting information because people feel comfortable with it and it allows for flexible approaches, which can be of aid if any issue were to emerge in the conversation (see Appendix D and E).

The reason behind the selection of this type of interview for this study lies indeed in its flexibility. Such flexibility allowed the interviewers to adapt the course of it depending on the development of the interview, taking into account the answers the participants gave them and the specific context of the institutions in which they used the rubrics. Therefore, it was considered that an interview would be the most adequate tool to measure the teachers' formation and the resources they have available to develop rubrics. Through the use of this tool, we could guide the teachers in order to get the necessary data from them as it allowed the researchers and interviewers to ask questions that needed more development. This is an advantage over other tools such as, for example, a yes/no questionnaire, where questions are not dynamic and there is no possibility to develop a course of data collection if answers require more explanation.

The interview was designed as follows: firstly, questions were intuitively created to understand both the process of design and application of the rubrics used by the teachers, the resources they had available, and the context (background) in which their students are in. Secondly, questions and responses where organized according to the research questions. Questions about the teacher's formation were added, since it is one of the factors we need to consider when analyzing the overall quality of the rubrics. Finally, the order of the questions was organized following recommendations from Dörnyei (2007) to conduct it in the most appropriate way. Accordingly, the first few questions were personal or factual questions, so that the interviewed felt competent and relaxed to answer the content questions. Therefore, content questions were placed in the middle of the interview. These were related to their experience with rubrics and resources and about their competences.

The interview was piloted in the institution Liceo N°1 Javiera Carrera. As indicated by Dörnyei (2007), a good interview requires piloting as it ensures that the designed questions can elicit sufficiently rich data, and that these questions do not dominate the flow of the conversation. Furthermore, piloting instruments is one of the strategies that can be put into practice in order to reduce measurement errors, both random and systematic. (see section 2.4.2). Thus, this pilot was conducted with an English teacher of the mentioned institution. It was recorded using an iPhone 6s plus, specifically the app 'Voice Memos' installed in the phone, which records voice audio format files. After this step, the data was analyzed via hearing the files and taking notes by the research team. Once the piloting and the analyses were completed, we were able to discern which of the questions we had previously made needed more development, such as the ones regarding the piloting of the rubric or the one alluding the statistical processes to guarantee validity. Moreover, the analysis of the interview also allowed us to be sure about the way in which we were going to ask these questions was the appropriate for the teachers to understand them (see the questions of the interview in Appendix in chapter 07).

On the one hand, one of the institutions used for the interview was the italian private school Scuola Italiana Vittorio Montiglio. This was the school to which one of the investigators went for all her primary education. The investigator contacted the English teacher and head of the English department and explained the project to her. The teacher pointed out our intentions to the director of the school, who gave us the permission to proceed. The investigator, then, through the contact of the Head of the Department, arranged an interview with the two English teachers in charge of *segundo medio.* The interview was conducted in one of the offices of the school by the investigator that provided the contact and another investigator of the team.

On the other hand, the second institution that participated in this investigation was the public school SIP (*Sociedad de Instrucción Primaria*) Escuela Arturo Toro Amor to which one of the investigators went for seven years. The investigator contacted the school first through the Director who approved the project and then contacted the head of the English department. She agreed to proceed with our project and gave us all the information about the teacher in charge of the level that we wanted to work with (*segundo medio).* After the meeting, we contacted the teacher and we scheduled the interview that was carried out by the same investigator that contacted the school.

In both cases the interviews were recorded and transliterated. For the first school Scuola Italiana Vittorio Montiglio, we used a smartphone Samsung A50 to record the interview with a voice record application from the same smartphone. For the second school Escuela Arturo Toro Amor, we used an IPhone 6s Plus to record the interview with 'Voice Memos' from the same smartphone. In order to transliterate both interviews and the pilot interview, we uploaded the recordings to a Google Drive folder where all investigators had access to and then we transliterate the recordings to a Microsoft Word file (see full transcripts in Appendix C and D)

3.2.2 Rubrics and documents

In this section, the rubrics and documents, namely the programs of the courses given by the teachers, will be explained. In the case of T1, two rubrics were given to be analyzed. In the case of T2 and T3 two rubrics were given to us as well. Some of the rubrics presented by the teachers present similarities between them. More specifically, two of the rubrics (one from T1 and the other from T2 and T3) deal with an oral exam, while the other two deal with an oral presentation and the making of a video in which the students have to cook a dish.

3.3 Data analysis procedure

The analysis of the study was carried out by using the content of the answers to the semi-structured interview made, the study programs, and the rubrics that we were given from the teachers from each establishment. The data was analyzed through a qualitative method.

From the interview, research questions were grouped in three different sections, one for each of the first three research questions of this study. Research question 4 was analyzed once all the previous questions had been responded.

This section will present the meta-rubric we used in order to evaluate the different aspects rubrics, according to us, should possess. It must be mentioned that the meta-rubric is not an instrument to collect data, but to analyze it. Later, the data analysis procedure regarding each research question will be addressed.

3.3.1 Meta-Rubric

A meta-rubric was designed to analyze and evaluate the rubrics used in the institutions. This meta-rubric describes in different sublevels the dimensions, descriptors and levels of the rubric under analysis. The tool was designed to help us to qualify the rubric in terms of its validity and reliability.

Meta-rubrics are described by Stevens and Levi (2004) as "a rubric used to evaluate rubrics. [...] It helps refine and polish some of the details in the rubric." Following this line, we used our meta-rubric that reflected the aspects of a rubric that are essential for any rubric to possess in terms of its design and application procedures (see section 2.2 of rubrics in Theoretical Framework in chapter 02).

In order to come about this meta-rubric, the first step that was taken was to recollect information of previous studies about rubrics and their components. Among those studies, we reviewed works of Fulcher (2010), Brown (2004), Hamidi (2010), North (1995), and

Jonsson and Svingby (2007). Subsequently we proceeded to summarise the essential categories that a rubric should possess, in which way these rubrics were designed, and what functions they had. We also examined rubrics that were used in our own undergraduate program in prior years to analyze and discuss their components and our impressions about them.

The meta-rubric designed provided therefore evaluation criteria of standard components of rubrics, including levels, dimensions, sub-dimensions, and descriptors. The information in the rubric was categorised into the aspects of construct and procedure (see the complete meta-rubric in Appendix I in Chapter 7).

3.3.2 Research question 1: Which is the formation and knowledge that teachers of English as a second language have at the moment of developing and evaluating rubrics?

To answer the first research question, the answers to questions 9 and 10 of the interview were analyzed in order to identify the competence and knowledge the teachers had respecting evaluation and rubric design. The questions were aimed at asking the teachers about their previous knowledge, if any, about the design and application of rubrics. Moreover, if their answer was, if fact, that they had received previous instruction of any kind, they were asked in which circumstances and what kind of instruction they had received.

We characterized participants as more or less prepared in their knowledge of design, development, application, and evaluation of rubrics. Furthermore, we made a differentiation between those who had received instruction of any kind by the institution they work for or if they were self-trained (i.e., they took a course or did their own research about the topic). In doing this characterization, we were able to discern whether the teachers' previous knowledge and formation limited or restricted the way in which they developed and applied their rubrics.

3.3.3 Research question 2: What are the institutional resources that the teachers can use for the design and application of the rubrics?

For this research question, the tools used were the transcription of the responses that the teachers gave when they were asked the questions 1 to 8, in relation to the resources the school provided for them to design and apply rubrics. We then proceeded to read each response to those questions and grouped them in the groups we will describe below. Afterwards, we characterized the resources of the teachers according to our criteria analysis.

We grouped the resources that the teachers mentioned into three different groups. The first group was “Training and further education” and it consisted on all the courses, workshops, and classes they had taken to learn more about evaluation and, specifically, rubrics. It also included the access to research sources regarding evaluation and rubrics, both in paper and through the Internet.

The second group was “Academic load”, including data about the hours that the teachers had to work on rubrics, the quantity of students per teacher, and also the quantity of different processes of evaluation that the teachers need to work on. This distinction was made because it is not the same to have thirty students from the same evaluation process than having ten from first year, ten from second year, and ten from third year. In simple words, the number of students in every class is significant in terms of academic load since the less students per class, the more efficient the class would be, and the less the time used in them. Due to that, teachers have to develop different tools, tests, and processes of evaluation for each course depending on these factors.

The third of resources was “Infrastructure and equipment”, which included the physical materials that the teachers could use to develop and apply their rubrics, such as pedagogical materials (textbooks, glossaries, etc.), physical space, and technological equipment such as a computers, tablets, and Internet. We divided the resources into these three groups for a better management of the data we were to collect, and also because they fulfil different purposes for the teachers’ work.

Afterwards, the information gathered from the responses was analyzed according to whether the resources were sufficient to develop the rubrics or not. This criterion was discussed among the researchers and we came to an agreement to value the different type of resources firstly as direct or indirect, depending on how they affect the quality of the rubric, secondly into whether they allowed/ prevented or limited/broadened the way in which the final rubric was made.

The questions regarding the type of resources were crucial to finally identify the nature of the resources that affected the making and application of the rubric. We interpreted these results as the different factors regarding resources that affected the rubric, separating them in terms of how critical and important they resulted to be for the teachers at the time to building, modifying, and applying the rubric. Thus, they were separated into the time that the teachers disposed in terms of how many students were under their charge, the number of classes, and the planification hours each teacher had. Moreover, we also interpreted the results in terms of the way in which the teachers worked, that is, if they did it alone or if they had a partner or a superior that would give them feedback on the rubrics they created. We also deemed as limiting or not that the three teachers received rubrics from the institution that could not be changed in any way.

As for the next factor, we interpreted infrastructure in terms of its importance over the final result of the rubric, where a base ground was needed, i.e., there were aspects of the infrastructure that were key to the quality of the rubric, while others were not as important or critical.

3.3.4 Research question 3: Which is the way in which teachers design and apply the evaluating rubrics to their students?

In order to answer the third research question of this study, we used questions 11 through 26 of our interview (See chapter 7 of Appendix). These questions were useful to give us information of how the teachers were able to come up with their rubric, how they created it or if they had adapted it; how they did this and from where they had obtained the rubric; how many rubrics they use; the people who were responsible in this process; how they guaranteed the validity of their rubrics, and whether or not their rubrics were aligned with the task the teachers were evaluating. In addition to this, there were also questions regarding the piloting of the rubric from the teachers, as well as the feedback they provided their students.

The answers the teachers gave us served as a means of having a clearer understanding of the processes that they go through when elaborating, modifying, and applying their rubrics, while also shedding light on the way in which the teachers provided feedback to their students. Furthermore, having this information allowed us to designate, with our meta-rubric, how adequate the preparation of the rubrics was and, in turn, their adequateness provided us with their overall validity. In order to do so, the analysis made with the meta-rubric in terms of procedures aimed at describing the designing process of the four rubrics the teachers of the two establishments provided us with as well as information on the feedback given by them to the students, the way in which they calibrated their work, and how they assigned a final score. All of these aspects were analyzed with the basis of the interview made to each of the teachers. Furthermore, the different elements of the rubric, meaning dimensions, levels, and descriptors, were also taken into account, since in some rubrics the teachers claimed to have modified certain aspects.

3.3.5 What is the relation that exists between the quality of the rubric and the interaction between the formation and knowledge of the teachers, and the institutional resources in relation to construct validity and reliability?

In order to analyze this question, the first three questions of the study had to be answered and analyzed first. As questions one, two, and three were being analyzed, we found connections to the fourth question. These connections were in terms of the interaction that the formation and knowledge of the teachers had about rubrics in relation to the institutional resources provided by each of the establishments in terms of construct validity and reliability. As these connections were found, we began a parallel answering of the first three questions as well as the fourth one. Triangulating the results found in the questions one, two, and three allowed us to make an in-depth analysis of our fourth question. Along the way we found out that it was crucial for us to characterize the resources given to the teachers and their formation as either limiting or restricting, which in turn allowed us to describe the way in which these two aspects affected the quality of the final rubrics. (see Results in Chapter 4 below)

Moreover, the analysis made of each of the rubrics the teachers provided us with served as a way to understand more clearly the way in which the teachers' competence affected the rubrics that they modified and applied to their courses. This, along with the information gathered in the interviews performed, was key to finally being able to make a connection between all of the aspects that were described beforehand.

In the last stage of the analysis made for this research question we began a discussion surrounding the elements that affected directly and indirectly the quality of the rubric. Those elements responded to two contexts, *human* and *non-human,* that provide us tools to characterize each case in terms of how was the rubric made, what were the parameters that teachers had when they adapt the rubrics given by each school, how many things they use in order to develop their rubrics, among other aspects.

Finally, we took every aspect regarding the resources given by each school and the previous formation of the three teachers interviewed, and we established the relations that they had in each case.

In the next chapter, we will present the results of this investigation that includes the answers per each research question based on the answers to the questions in the interviews conducted to the teachers and their corresponding summarizing table.

Chapter 4: Results

This chapter will present the results of the questions from the interviews. In order to present them, the information collected from the questions will be organized regarding each research question.

The chapter is organized according to the order of the research questions established on Chapter 3. First, we will be addressing the results regarding the formation and knowledge that teachers have at the moment of developing and evaluating rubrics. Secondly, we will present the results for the institutional resources that teachers have in order to develop the rubric. Later, we will deal with the information collected about the way in which teachers design and apply their rubrics. Finally, there will be a brief explanation about how we discuss the results for the interaction between the formation of teachers about rubrics and the institutional resources and how those variables affect rubrics.

In addition to the information described after each research question, at the end of sections 4.1, 4.2 and 4.3, there will be a summary table due to present the results in a much condensed way.

4.1 Research question 1: Which is the formation and knowledge that teachers of English as a second language have at the moment of developing and evaluating rubrics? (questions 9 and 10)

On the one hand, T1 reports that she does not have previous knowledge about rubrics, their design, or application. However, she does mention that she has some basic knowledge about rubrics from her methodology class at University. T1 also reports that she looks up for rubrics on the internet and then adapts them.

T2 reports she had formation of rubrics from University and from an Institute where she gave PET tests. Here she was trained with Cambridge rubrics where she revised and discussed the rubrics in order to learn how to use and apply them. She was also trained in terms of calibration because they had to evaluate in the same line as Cambridge standard evaluations. She gave an example where she had to watch a video of a student and then evaluate his or her performance, comparing also her results with the ones given by the institute.

Finally, T3 also had formation on rubrics at University and from a professional institute where she was trained with the Institute’s rubrics. They revise those rubrics and discuss them in order to learn how to use them. In the process they calibrate their evaluations with the standard evaluations given by the Institute. They also made them criticize and question their decisions regarding the process of evaluation with the rubric.

Table 1 *Summary of results from research question 1*

RQ1	Institution 1	Institution 2	
Informers	T1	T2	T3
Did you have previous knowledge of rubrics?	No	Yes	Yes
Do you have formation of rubrics?	Yes	Yes	Yes
What kind of formation on rubrics do you have?	University	University and Training at Institute	University and Training at Institute

4.2 Research question 2: What are the institutional resources that the teachers can use for the design and application of the rubrics? (questions 1 to 8)

T1 reports she is in charge of 9 classes, distributed in 5 different levels, having 40 students per class approximately. In order to design the rubrics for these classes, the school provides the following resources: a basic teachers' room and a Chromebook laptop (used to work in class and to design rubrics). In addition to these resources, T1 is paid one hour for planning, from which only 20 to 40 minutes are used for rubric design. Were that not enough, she has to work on the weekend as well, not only to design rubrics but also to look for different rubrics on the internet. Regarding T1's training, she reports that the institution has not supplied courses regarding rubrics or evaluation. Moreover, she claims that the institution's training is sparse, has little to no variety in terms of topics, and no particular focus on evaluation.

T2 and T3, who belong to the same institution, report that they are provided with computers, laptops, projectors, classrooms with a good sound system and Wi-Fi. Moreover, they dispose of an office exclusively for the English department (supplied with computers, Wi-Fi and desks), they are also given free access to both the school library and photocopier/printer. Furthermore, T2 and T3 claim that the institution is equipped with an academic platform (School Track) which they use to share material with students and other teachers.

Regarding their classes, both T2 and T3 are in charge of 6 classes each (having 20 to 25 students per class). T2's classes are distributed in 5 different levels (*quinto básico, primero medio, segundo medio, tercero medio* and *cuarto medio*) whereas T3's classes are distributed in 3 different levels (*segundo medio, tercero medio* and *cuarto medio*). In terms of their paid planning hours, both T2 and T3 indicate that they dispose of 10 hours in total, reporting that the school provides 40% of their teaching hours for planning. However, they only use 2 to 3 hours for rubrics. Moreover, T2 claims that she spends different hours and days working separately on the same rubrics, while T3 declares that she uses Wednesdays exclusively for working on rubrics, as she has no classes that day.

Table 2 *Summary of results from research question 2*

RQ2	Institution 1	Institution 2	
Informers	T1	T2	T3
How many classes do you have?	9	6	6
How many students per class do you have?	40 approx.	20 to 25	20 to 25
How many levels?	5	5	3
How many paid planning hours do you have?	1	10	10
How much time do you spend on the elaboration of rubrics?	20 to 40 minutes	2 to 3 hours	2 to 3 hours

4.3 Research question 3: Which is the way in which teachers design and apply the evaluating rubrics to their students? (questions 11 to 26)

T1 explained that some of the rubrics that she applied are provided by the school itself. Others were designed and modified by her, prior to test application, using pre-existing rubrics she finds on the internet as a reference. The rubrics she is given by the school were designed prior to the task evaluated, years ago, by their superiors. T1 modifies the rubrics by herself with the exception of the rubric for oral exams, by changing or adding what she thinks is necessary. If she considers there is a dimension that is missing in the rubric, she intuitively knows that she has to modify and add such dimension. Furthermore, she argues that her rubrics are aligned with the tasks she is to evaluate, since she always ensures that the rubrics she modifies suit whatever specific task she is asking for. On that same line, she uses the rubrics in order to evaluate the students and grade them, and if she finds that there are aspects that are missing or that are irrelevant, she modifies the rubrics for the next year for the coming courses.

Since T1 explained she had not had any prior formation or training regarding evaluation and the use of rubrics, it is interesting to note that she is aware that rubrics are useful to evaluate and score a test in the fairest way possible. This suggests that the teacher probably has a notion about the concept of fairness, which could arguably be framed as part of her expert judgement as an alternative to the lacking theoretical and technical support, i.e. the necessary formal articulation based on a previous training. Finally, in terms of applying the rubric, T1 explains that the rubrics from each class are fairly similar to each other and thus she makes sure to explain the rubrics beforehand to her students. In that way, they are aware of the way in which they are going to be evaluated considering their corresponding levels.

As for feedback, T1 mentions that she used to give feedback to each student separately at the end of the class, where she pointed out minor mistakes and errors. She would also ask the same students what they thought their most common mistakes were, and then she would write them down on the whiteboard. When asked about the way in which she gave feedback again, she responded that she did it orally, to everyone in the class, because according to her that was the most effective way of giving feedback to her students, since then everyone could learn what they were doing wrong or what they were doing correctly. This kind of feedback is given to the students each time they have to make an oral presentation (afterwards).

In the case of T2 and T3, a different type of feedback is given depending on whether the evaluation is oral or written. For oral evaluations, they give feedback individually at the end of the oral exam along with the corresponding grade. In this kind of feedback, they discuss the weak points and strengths of the student. As for written evaluations, they give written feedback in which they use symbols that the students have already learned beforehand (for instance, spelling would translate to ‘sp’, and word order to ‘wo’). After every evaluation, both T2 and T3 also give individual feedback to make sure to answer any questions the students have and to clarify doubts.

On the other hand, T2 and T3 report that they always work together for *segundo medio*, as they think that the same rubric should be applied to the same level of students. Therefore, they create and adapt every rubric together on their planning hours in order for them to be applicable to the students of both teachers. After adapting and changing their rubrics, they have to present them to the head of the English department for her approval. Only when she approves the rubrics shown by the teachers, they are able to use them. They also mention that they cannot use the same rubrics for different years as needs and students change. Therefore, they re-evaluate every rubric and change it before applying it again.

T2 and T3 take rubrics from FCE (Cambridge First Certificate in English), the ones used at school on *cuarto medio* for *maturitá*, an interdisciplinary exam in which students write about a random topic from their classes of their last year. These rubrics borrowed from the *maturitá* are created by the Italian commission of teachers in Italy. The task is evaluated by their teachers and an Italian commission of teachers. After they pass the exam, students are given a certificate that confirms they have an education which is equivalent to the schools in the Italian national system. T2 and T3 also search for rubrics on the internet in general. The ones they use with no modifications are the ones for the *maturitá* and FCE rubrics for speaking and writing. For the rest of the rubrics they apply, they use internet as a reference and adapt the rubrics they find there to the rubric they make in order to resonate to the task, the evaluation, and what they can demand from the level of their students.

T2 and T3 work together to create or adapt the rubrics they use at school. They take a rubric from FCE, for example, and they analyze and adapt it according to their needs and their students. They mention the example of an FCE rubric that expects the examinee to use passive voice and perfect tenses, in which case the teachers change that requirement to the level of students of *segundo medio.* Also, for a video task, they use FCE rubrics but change their descriptors as they identify levels of the rubrics in which the requirements do not match (e.g. some dimension is evaluated and descripted in the highest level but not in the lowest one. Apart from the FCE, they borrow rubrics from the maturitá of *cuarto medio* at school. T2 and T3 find these rubrics exhaustive and complete. However, they adapt their expectations to students from *segundo medio* and not *cuarto medio*, without changing the rubric itself but their own expectations towards the students.

T2 and T3 also argue they use their rubrics to evaluate processes and products. For evaluations that need research and that involve a process, they include the research process and classwork on their evaluation and, consequently, in their rubrics (e.g. a video project). In the case of writing and speaking tests they only evaluate the product, although they do practice in class but without evaluating that stage.

Regarding statistical processes to ensure reliability, both T2 and T3 claim not to do any of them in terms of actual mathematical calculations or formal reports. However, they do have conversations in regard to, for instance, the degree of importance of each dimension for their rubrics, or the difficulty of the items applied to previous tests. This indicates that working together constantly is their way of ensuring reliability without using alternative statistical processes, as they do not have time to work on them.

Their rubrics are, according to them, always aligned with the task they are evaluating. Identifying the task is the previous step before starting to create/adapt their rubrics. The teachers do not pilot their evaluations before applying them, as they claim not to have time to pilot. However, they also know what to expect from students because they know them very well and they also use their previous experiences on the same students to improve their next evaluations.

Instead of piloting rubrics, T2 and T3 also practice the forthcoming evaluations with their students. As an example, they gave a test in which students had to re-organize given information in order to write a biography of an artist. Some days before the test, students were asked to do the same task but with a different artist from the one they had to write about on the test. The teachers put considerable emphasis on the way in which they practice for the tests with students, reporting that they do not reveal the fact that it is a practice for a test, but rather use it as a part of the learning process. The purpose here is that students do not to strive only the day of the test but they also work hard during the other lessons. In this sense, T2 and T3, although not piloting their evaluations, seem to be aware of the need for students to practice tasks before actual testing.

Regarding feedback, T2 and T3 explain that their practices depend on the type of evaluation. For evaluations such as oral interrogations, they give feedback to the students as the exam has finished. Students also receive their grade immediately after the evaluation. The teachers meet then with the students individually and they explain the shortcomings as well as the strengths observed in their performance. They argue that the most effective way of giving feedback is in a “one by one” modality, because the perception of the student about his or her own work changes. According to the teachers, the students are given the space to understand their performance and the areas in which they need to pay more attention and, what is more, they frequently show their appreciation of the comments on weaknesses they can further work on. The teachers claim that the students value this modality, and argue that it must be done that way, because a general feedback is not feasible when evaluating such personal answers. Moreover, this practice is helpful when creating a bond between teacher and student.

In the case of written evaluations, the teachers write editing symbols which the students know beforehand, since *primero medio*. For instance, they underline mistakes such as “spelling” and “word order”, and write the symbols “sp” and “wo”, respectively. The students are expected to recognize these symbols, understand where they failed, and be aware that, in the end, these annotations imply a lower grade. After handing in the revised test, the teachers also meet in the aforementioned “one by one” modality to check if they have any questions and not only to clarify their doubts, but also to encourage them in a positive way for the next evaluation.

Finally, when it comes to writing evaluations, teachers use to pay attention on grammar, spelling, punctuation, register, paragraphs, and all formal aspects. Nonetheless, they recognize that for open questions as in the ones analyzing books, those formal aspects, although important, are not the essential aspects to evaluate. In those cases what is more important for teachers is how they organize the ideas and the content, and how they defend their perceptions or interpretations on the answer to the questions. Teachers also mention that the rubrics usually exhibit the importance that is given for each dimension according to the evaluation.

Table 3 *Summary of results from research question 3*

RQ3	Institution 1	Institution 2	
Informers	T1	T2	T3
How do you do the rubric?	Adaptation and use of the rubrics given by the school	Adaptation and use of the rubrics given by the school	Adaptation and use of the rubrics given by the school
From where do you take the rubrics you use?	Internet and from the school	Internet and from the school	Internet and from the school
How many people participate on the creation of the rubric?	1	2	2
Do you pilot your evaluations?	No	No	No
Do you show the rubric to your students before the evaluation?	Yes	Yes	Yes
Do you give feedback?	Yes	Yes	Yes
How do you give feedback?	In general, after each class where students present, to the entire class orally	In general, after the tests and one on one	In general, after the tests and one on one

4.4 Research question 4: What is the relation that exists between the quality of the rubric and the interaction between the formation and knowledge of the teachers and the institutional resources in relation to construct validity and reliability?

In order to answer research question 4, we had to make an in-depth analysis with the results from research questions 1, 2, and 3, since this question focuses on the relationship between two variables that affect the rubric. Therefore, the results for this research question will be presented in Chapter 5, Discussion of Results (Section 5.4) To sum up, interactions between formation and resources were observed and they worked in an interdependent relationship. In some cases, formation affected quality regardless of resources (as in the case of formation affecting expert judgement, and that leading to teachers autonomy for making changes in the rubrics), while in others resources affected quality regardless of formation of teachers (as in the case of received rubrics from the institutions). Within resources time seemed to be the most important.

4.5 Rubrics from Institution 1: Escuela Arturo Toro Amor

Two rubrics we were provided from this school will be presented. The first one, "Talking Point", describes an oral presentation that the students had to present. The second rubric, on the other hand, refers to an oral exam. To see both of these rubrics, see Appendix F.

4.5.1 "Talking point"

At a first glance, it can be seen that there are four sub-dimensions (see Appendix F) . There is no presence of levels, and the descriptors (that serve more has sub-dimensions) are not accurate. A mixture between descriptors and sub-dimensions can be observed in this rubric, and that is why mainly descriptors are not clear. Also, only the number of maximal score is stated, so it is not very clear what should happen when a lower level performance is observed or the specific score to use in that case. The aspects in which the score points can be better understood in how they correspond to every performance are the aspects added by the teacher. However, even in that case, the alignment between what is being evaluated and the total score that allows evaluators to assign a fair score is not clear.

In general terms, the rubric included assessment of classwork aspects as a dimension, thus mixing the evaluation of process and results in the same descriptors. Using the time appropriately during class is something that responds to the process being evaluated, not the result, and it is not clear if the task as a whole is also evaluating the process. This sub-dimension was added by the teacher. On the other hand, it is also not clear how the "attention/respect between students" aspect (point 2) could be evaluated because paying attention and respecting others students could have several meanings. For example, it could be about not talking during presentations or about answering questions, etc. Either way, what this sub-dimension means is not stated anywhere in the rubric and, therefore, it is dependent on the judgement of the teacher and thus it loses reliability, since there is no systematicity in the way in which T1 scores the rubrics. Furthermore, since there is no systematicity and much of the score assigned relies on her judgement, the consistency in her scoring would be affected.

As for the Language dimension, none of the sub-dimensions are described, hence the information is vague. Although vocabulary and grammar are mentioned, the rubric does not specify what kind of vocabulary or grammar is evaluated. In terms of the aspects of construct validity (see Section 2.3.1), this serves as a problem for the content and structural aspects, since there seems to be no link to what the program of the course is aiming for and what is asked for in the rubric. In other words, there is no correlation between the construct of the task and the program of the course. This would serve as an issue in terms of the pedagogical aspect of our meta-rubric. Furthermore, it can be seen that all four aspects of this dimension correspond more to the category of dimensions than descriptors.

Also in this Language dimension, the description is not accurate in point number 3, since it is about clear and adequate pronunciation, but it is not specified what adequate pronunciation refers to. In other words, the quality of adequateness is not established, although it can be assumed that it would have to be adequate in terms of what the program says. This goes hand in hand with the description "speaking fluently", since the same problem can be perceived.

There is also more emphasis and valoration (i.e. higher scores) in pronunciation and fluency rather than the language used (grammar, vocabulary), which is why the name of this dimension is not accurate, and according to the meta-rubric it affects construct validity.

In the case of Format, point one, regarding time, seems to belong to the format, but it does not seem to be totally accurate since the range 1:30 to 3:00 minutes is not justified, and it is not very clear how and why the score would be lowered from the two points maximum. This may constitute a problem for the validity of the task, since it means that a student who speaks 3 minutes will be scored just as one who only spoke a minute and a half. In addition to this, it also affects reliability as it makes it difficult to know which score corresponds to an observed performance. If the range between minutes was smaller, this problem would be partially solved. This unclear sub-dimension was added by the teacher, which seems to reflect poor decision-making that the teacher has due to her lack of knowledge on rubrics (as discussed in Section 5.1).

It is not entirely clear why the second point, regarding the sending of the visual material to the teacher's mail, would belong to the format, just as it is not clear how points can be assigned, since the maximum score points is 2. This sub-dimension seems more coherent with "time management" rather than format. Point three of the rubric, regarding the use of voice, does not seem to be related to format at all and it seems to be be a separate sub-dimension as well as the second point (visual material).

Finally, with respect to Content, point one, regarding the organization of information is not strictly related to the content, but rather to the organization of the task. In view of its content, this sub-dimension seems to be a separate dimension on its own. In this same descriptor, restrictions to the use of translators or copy-paste are included, but it is not clear why this belongs in the same point as the organization of ideas and not to the format dimension. Furthermore, it seems unfeasible to measure if the students use or not these tools. It is possible that the teacher is aware of the way in which each student speaks, but this seems unlikely under the circumstances. It could be possible to observe if students stick to the vocabulary and grammatical structures that they have been taught, but this is not explicitly told in the rubric.

The descriptor that refers to the questions and answers is the most accurate in the whole rubric, since it assigns a score of one point for each question (2), and it is understood immediately that if one of the answers is wrong, the score is lowered. It is interesting to mention that this descriptor was added by the teacher, given that the rubric is given to her by the school. Since T1 does not appear to have previous formation and it has been said before that her decisions are not the best due to the lack of expert judgement, the fact that this descriptor is the most accurate can be explained in terms of her expert judgement, because she had done it before successfully or plainly due to her experience-based common sense as a teacher (see Section 5.1)

4.5.2 Oral exam rubric

The second rubric from institution 1 chosen to be analyzed was an oral exam rubric. This rubric included four dimensions: understanding, use of English, pronunciation, and communication of ideas. Each of these dimensions includes with three descriptors, and every descriptor gives some explanation as to what every dimension covers. There are six levels portrayed as percentages: 100% to 90%, 89% to 85%, 84% to 70%, 69% to 45%, 44% to 31%, and 30% to 10%, as well as 6 bands, from 0 to 5. Each band corresponds to a different percentage, 0 the lowest, and 5 the highest. This oral exam rubric is externally mandated (see Appendix G) since it is not the teacher who designs it, but rather the institutions' stakeholders, and the rubric is used from 1st to 12th grade (*1° básico* to *4° medio*).

In the first place, the use of percentages affects the rubric's reliability, due to its excessive variation in its results. By having such diverse percentages, the rubric's design makes it harder for T1 to assign a proper score. In fact, T1 says that she does not pay much attention to the percentages present in the rubric (see Appendix D) for the complete interview).

Secondly, the range of the percentages is not reliable since there is too much difference between them and that may alter the assignation of the final score. For instance, there is a range from 69% to 45% in the second band of the rubric. This implies that a person that gets 69% will be scored the same amount of points as the person who got 45%. This means that it is not completely clear how to reach intermediate percentages, since it seems that students that are scored in this band get the same amount of points notwithstanding the percentage they got. For instance, in the case of the range 69% to 45% it is not specified how to reach 50%. This is a problem for the reliability of the rubric (scoring aspect) in the sense that it is left to the criterion of the teacher to decide what the difference between 69% to 45% is. Finally, the rubric lacks a correlation between the performance and scoring results. That is to say, if a student obtains 90%, he or she would get the same final score that someone who got 100%, and so on with every range of percentages. This affects reliability because the difference between 90% and 100% is left to the criterion of the teacher, which may vary.

In terms of the first dimension, named "understanding", the second and third descriptors seem too vague. There are concepts as "some" or "considerable" that should be explained with precision, so stakeholders know which levels those concepts are referring to. In addition, the first level of this dimension asks for "absence of repetitions", repetitions that are necessary in any level. The problem with this is that there might be external factors that affect the way in which the student understands the idea he or she is being presented with. For instance, that the question/prompt is not clear itself, hearing issues that the student may have, noises coming from external sources, etc. This may affect construct validity because asking "absence of repetitions" is not incorporated on the program (see Appendix H and I), therefore, the task is not aligned with the content of the program.

The first dimension also includes ideas that are too unlikely and unclear. Firstly, because there may be many external reasons which could affect the result of the rubric (wording, external noise, etc.). Secondly, concepts as repetition and prompting are not self explanatory. Because of this, we gather that it is very likely that the students will not be able to understand this dimension without further explanations of both the rubric and the task they have to carry out. However, since there were no follow-up questions in the interview of T1 involving the nature of this evaluation, we cannot know for certain what the task of this evaluation actually is, since there is also no trace of it in the program of the course that T1 handed to us.

In the second dimension, regarding the "use of English", the construct is not clear and neither is what the teacher means by expecting to have a good control of simple grammar. In the first descriptor, it is implicit that elementary mistakes are not allowed or will not be made because, in the following descriptor, a few elementary mistakes are made or allowed (allowed in the sense that the presence of them are penalized). In terms of format, two out of the three descriptors have two descriptions, while only one possesses three. At level one and two it is intended that there is appropriate vocabulary for the level, however, it is not entirely clear how the quality of appropriate is observed.

At level three, the use of grammatical structures and the use of vocabulary are together at the same point within the descriptor, while the others, vocabulary and grammar were separated at different points. ‘Few’ and ‘many’ vocabulary mistakes seem to be vague indicators as it is not clear how much ‘few’ and ‘many’ is.

In the third dimension, pronunciation, the first descriptor includes the concept "intelligible" described in all three levels, that is, the same is expected in all of them. There is no variety and it is not clear how the score would be lowered if intelligibility is not accomplished. Furthermore, good control of sounds and stress is required. It is not clear what “good control” means, and there is no reason why stress is to be evaluated, since this is not clearly stated in the program of the course. Additionally, in the last two levels the descriptors are the same. This is a problem due to the fact that there would be no distinction between bands 1 and 3, which makes the assignation of a percentage more difficult.

In the final dimension, which is called “Communication of ideas”, descriptors show the highest level of consistency and clarity in relation to the descriptors from the other dimensions. Every descriptor of every level is clearly related to the corresponding dimension, due to the existence of continuity, meaning that each descriptor is a follow-up of the other. For instance, in the first level, answering questions are expected according to the level, meanwhile, in the following levels, descriptors also allude to answering questions to a different degree of requirement. There is still a use of vague adjectives such as ‘some’ and ‘considerable’.

4.6 Rubrics from Institution 2: Scuola Italiana Vittorio Montiglio

Two rubrics we were provided from this school will be presented. The first one, “Simulazione Colloquio Inglese”, describes a final oral exam that the students have to present. Meanwhile, the second rubric is a video project named “Fabulous Food”. To see these rubrics, see Appendix E.

4.6.1 Simulazione Colloquio Inglese (oral)

This rubric included three main dimensions, each one with a different descriptor and score. The first dimension regards general and specific knowledge of the topics studied in classes. This knowledge is evaluated by different qualifying descriptors, which ascend depending on the performance of the student. Thus, the lowest level is “incomplete”, followed by “fragmentary”, “basic”, “satisfactory”, “adequate”, “complete” and “precise” and “extensive” and “profound”. It is worth mentioning that this dimension presents no exhaustive description of each level, which is why it is not entirely clear what each of its valuations mean specifically. This is interesting, since the rest of the dimensions do present levels with their respective descriptors describing each level at length.

The second dimension refers to processing, logical, and critical capabilities. Here, the descriptors of each level describe, in a continuum, the way in which the student synthesizes and re-elaborates the contents in an autonomous and original way, in a personal way, to their own knowledge, exposing the contents offering personal ideas, in a linear way, and finally without making any synthesis. The dimension presents, therefore, many levels, which suggests thorough thinking and assessing expertise on the part of the designers. However, some of its descriptors are not totally clear at first glance, for instance, the second descriptor makes reference to the synthesis of contents in a personal way, but at the same time, there is no explanation of what is being referred with “personal way”.

The third dimension refers to the properties of “expressive instruments”, in other words, expressiveness of the oral presentation. This ranges from the lowest level, that is, lacking expression and communicating with difficulties, to the highest level, which is expressing oneself clearly and correctly. Like the former dimension, this one presents continuity in its descriptors. Unlike the former, these descriptors describe very clearly what is expected from the student to get a certain score.

One aspect worth highlighting is the presence of a conversion table, where the scores are converted to a grade. This allows the student to know beforehand how their evaluations are going to be scored, so they already know the score they should obtain in order to pass.

There are some elements of this rubric that do not offer any explanation, which is the case of the scoring of each level, which varies depending on the dimension. For instance, there is a level in the second and third dimension that assigns 5, 4, or 3 points for just one aspect. The descriptor, in the case of the second dimension, is also rather vague, since it is not clear how the scoring in this particular level is done. It is not clear how or what is expected from the student to get 3 points, and what the requirements are to obtain 4 or 5 points.

Overall, the descriptors found in this rubric are, for the most part, thoroughly explained. Notwithstanding one or two cases, these descriptors manage to describe very clearly each of the different levels of this rubric. Therefore, students know what is expected from them.

4.6.2 Project "Fabulous food" (video)

This evaluation uses three different rubrics: classwork rubric, video rubric, and speaking rubric. Initially, this implies a better organization regarding the tasks that are required from the students. Each of the different rubrics seem to be fully completed (descriptors, dimensions and levels are clearly distinguishable). Because they are fully completed, rubrics are easy to read and comprehend. The rubrics come together with the evaluation, therefore, the students have a detailed description of the task at hand. Moreover, it is known that the teachers present the rubrics to the class before the evaluation.

For the Classwork rubric, students are asked to behave appropriately in each class, to respect their classmates, and for them not to be necessary to be reminded what work to do, or to ask them to be silent. Another point is that they must work on the assigned tasks every day, handing in a glossary with new words and interesting ideas. Every time the students fail to meet the requirements, a 0.1 points discount is made.

As for the Fact File, there are three different dimensions: classwork/laboratory, which refers to the appropriate work during laboratory hours and to the non-waste of time (2 points), to the students handing back the fact file with new information each class (2 points), and to them sending the links of their sources to the teacher's email address (1 point). The second dimension is research, which refers to the students handing in the file with the required information, and with reliable sources in specific dates by the end of the lesson (1 point), and that the presentation of the fact file be clean and with clear handwriting (1 point). Finally, the dimension about punctuality refers to the video being sent to the teachers' email address on time (1 point).

In the case of the Video rubric, regarding "Narration/ acting/ character dialogue", there are three points (per descriptor) to evaluate by each level. The first one has to do with pace (good pace, good pace to match visual, rushes through or drags behind), but in the first level, the innovative use of narration/ dialogue is also evaluated. The second point refers to the use and variety of inflection, pace, and emotions, and the third point to the quality of the students' acting. In all of the descriptors there is a continuity in respect to the evaluated points.

As for "fact file and planning", perhaps the dimension's name is not quite clear because the fact file is also evaluated in the previous rubric. When reading the descriptors, it can be understood that it points at the information that was presented in the fact file and for this to be depicted clearly in the video. Just as in the previous dimension, there is continuity regarding what is evaluated (everything that is important is in the video, some things are depicted, does not exist).

In "Story/ message/ information", three points per level are evaluated: that the story and the message are clear/ a little bit clear/ unclear, that the story achieves its goal/ it is complete but does not achieve the goal/ it is not organized or not well expressed, and that the information is organized in a creative way/ it is coherent/ it is neither organized nor coherent. The last two points are somewhat similar because both of them refer to the organization of the test performed, besides, being creative and coherent has nothing to do with it, since it is referred to organization in the performance rather than creativity.

The Speaking rubric contains four different dimensions, and three different levels (5, 3, 1) separated by three other middle bands (4, 2, 0). Separating levels in this way allows the teacher to assign in-between scores whenever the student's performance shares features of two different bands. The rubric also includes a level where 0 score points are assigned whenever the performance is below band 1. Nevertheless, it is not totally clear when the performance should be below band 1, since this level is not described. It is assumed, therefore, that 0 points are assigned when the requirements for level 1 are not met.

In the first dimension, grammar and vocabulary, all three levels have a point in which a good degree of control is asked. The dimension shows a clear continuity in all its levels, since it also asks for some complex grammatical forms, an attempt of these forms, and none of these forms. The same happens with the next point, which is the usage of a range of appropriate vocabulary. This is required in all of the levels; the only thing that change is how the student is expected to give and exchange views on an (wide) range of familiar topics/ everyday situations.

In the dimension of "discourse management" we can see that there is no continuity in the different levels, since two of these present three points, and the last one only presents two (the point related to the contributions being relevant and repetition is not present). Other than that, the rest of the points do have continuity between them, except for the first point in the first level, where organization of ideas is expected. In turn, the second level describes that very little repetition is made instead of organization of ideas. A range of cohesive devices is also expected. This requirement was present in the program of the course, which in turn means that the students are acquainted with cohesion, a grammatical aspect that we did not find in the rubrics of the other establishment.

In the dimension regarding pronunciation, we can also see that there is no continuity in the points of each level, since the first two present four elements and the last one presents only one. This last level describes all of the elements of the other levels, but condenses them into just one point. The first two levels ask for intelligibility, whereas the last one requires only for the pronunciation to be mostly intelligible. The students are also asked for intonation to be appropriate, but it is not entirely clear what they mean by appropriate (i.e. in which terms it would be appropriate. It can be assumed that it is appropriate for their level, and if this is so, intonation should have to be something students had previously encountered). Sentence and word stress being accurately placed is something that is also required for both the first and second level (they are repeated).

Finally, the articulation of sounds is also evaluated. In the case of the first level, these should be articulated clearly, whereas in the second level these are generally articulated clearly. For the last level (1) some control of phonological features at both utterance and word level is evaluated. Since the level of English that the students are expected to possess is present in the program of the course, this does not constitute an issue, unlike what happened in the rubrics from the other institution (see Section 5.5.2). Due to this, the degree of control that the student has on grammar and vocabulary is to be understood in terms of the level of English of the students. Moreover, the fact that the students are asked to have control of certain phonological features should indicate that they have previous knowledge of this.

Lastly, in the dimension of interactive communication, there are two aspects (in the descriptors) in each level that are evaluated. In all of the levels it is expected that the student initiates and responds appropriately. in the case of the first level, it is also expected that they link contributions to those of other speakers. The next point refers to being able to maintain and develop the interaction with other speakers and negotiating towards an outcome. In the case of the second level, this is done with very little support, whereas in the last level the interaction is maintained with very little prompting and support. At first glance, it is not entirely clear what the rubric means with this.

4.7 Summary of results

Talking point: T1

In this rubric there were no levels. This decreases the rubric's reliability as it complicates the scoring process and it is left for the teacher to decide. Moreover, there is a dimension that is not related to the task itself, but to the process. This affects negatively the task-related validity of the rubric and also its reliability, as it is not systematic. In addition to this, the aspects that are being evaluated are not aligned to the program, which affects the pedagogical validity of the rubric. Finally, certainly an important focus is given to pronunciation and fluency and quite less focus to grammar and vocabulary. This has an impact on construct validity as it does not evaluate properly aspects that are relevant to the construct.

Oral exam rubric: T1

The fact that this rubric is designed to evaluate students from *primero básico* until *cuarto medio* is quite relevant in terms of the construct validity in the sense that these are different levels that you should not be evaluated in the same manner. Thus, we acknowledge that demanding the same level of knowledge from students from different levels is not appropriate as it is the teacher who has to adapt the expectations to its own criterion, which decreases reliability.

In the rubric there is a range of percentage that is certainly wide -and has no descriptors- (goes from 69% to 45%). Therefore, the difference between the 69% and 45% is left for the teacher to decide. This, in turn, decreases the reliability when it comes to scoring. Moreover, the descriptors of the rubric are certainly vague as it is not specified. We see descriptors such as "considerable", "some" and "good", which are also dependent on the teacher's criterion. This also decreases reliability. In addition to this, within the levels, there is one that says "absence of repetition". This, clearly, would not apply for students from *primero básico,* as this should not be applicable to all levels, which decreases reliability as well. This also decreases construct validity as they are not evaluating what has to be assessed. Furthermore, we found that some descriptors are repeated in different levels, which, in turn, decreases construct validity because each level should evaluate different performances. In addition to this, reliability is decreased as acknowledging the difference is left to the teacher's criterion.

Simulazione Colloquio Inglese: T2 and T3

Regarding this rubric we acknowledge that it is well aligned with the program of the course, which contributes to pedagogical validity. Moreover, the levels are well delimited and they have specific descriptor, which makes it easier for teachers to evaluate. However, there are some descriptors that are unclear (for instance saying *in a personal way*), then relying on the teachers criterion, thus decreasing the rubric's reliability. The vast majority of the dimensions are explained in a clear manner, contributing to reliability, as they allow teachers to agree on the same observations. In addition to this, they also have a table that converts the scores from the rubric to an actual grade, which assures that the process of scoring is reliable. The result is that it is quite explicit what is expected from the students, then resulting in a reliable rubric.

Project Fabulous Food: T2 and T3

When it comes to the rubric “Fabulous Food”, there are four unique rubrics that define and separate the tasks to be completed and evaluated. The corresponding rubrics are “classwork”, “fact file”, “video” and “speaking”. This is relevant as it ensures that the evaluation of each task is well established, hence, contributing to task-related validity. The fact that is does not mix the tasks in only one rubric, and also that it designs every rubric different from each other, contributes to reliability, then facilitating the evaluation for the teachers. Moreover, there is continuity between the levels. This contributes to addressing the levels of competence of the students, which contributes to both construct validity and reliability. Another factor that contributes to reliability is that the scores are well delimited. Therefore, the rubric does not leave much for the teachers to decide on their own, thus avoiding arbitrariness. Finally, the rubric is aligned with the program, thus contributing to pedagogical validity.

As a conclusion, we can say that the rubrics from both establishments seem to have several differences. These differences lie in the way in which the rubrics are designed, since the rubrics from T1’s school showed to have difficulties regarding the way in which dimensions and descriptors work, as well as showing in one case absence of levels. On the other hand, T2 and T3’s rubrics were of much better quality in terms of their design, and they seemed to respond in a better way to the program of the course that the teachers presented. Moreover, their rubrics are much clearer in terms of describing the task that the students are to be evaluated for. In some rubrics from both establishments, however, there seemed to be issues regarding the way in which the dimensions of the rubrics were scored, in that there were some instances where only one aspect was to be evaluated but three points maximum were to be assigned, making it a problem since it was not clear how or when less points should be assigned.

A discussion of these results is now presented in the following chapter.

Chapter 5: Discussion of results

The main objective of this study was to explore the relation between the formation of teachers, the resources they receive from institutions and the effects of this interaction in terms of the impact it has on the quality of the rubrics they use.

In this chapter, we will present the discussion of the results for our four research questions. The research questions were established as follows:

1. Which is the formation and knowledge that teachers of English as a second language have at the moment of developing and evaluating rubrics?
2. What are the institutional resources that teachers can use for the design and application of the rubrics?
3. Which is the way in which teachers design and apply the evaluating rubrics to their students?
4. What is the relation that exists between the quality of the rubric and the interaction between the formation and knowledge of the teachers and the institutional resources in relation to construct validity and reliability?

In order to go through the discussion of the study, we divided this chapter into four sections, one per research question. The following sections include a discussion of results regarding the formation of teachers regarding rubrics, how teachers design and apply rubrics, and the resources given by their educational institutions. Finally, we establish the relationships between both the formation of teachers and the resources of the schools that affect the quality of a rubric. To sum up, both teachers had a lack of formation regarding design and application of rubrics, however, they have developed and expert judgment (as predicted by Harris & Eggen, 1993), which helps them take better or worst decisions regarding their rubrics. We concluded that formation was a critical element in the design and application of rubrics and that affects directly to the quality of it in terms of construct validity and reliability. Resources affect mainly indirectly the quality of the rubric, and they may restrict or limit the process of developing and applying a rubric. From these, the one that seems to have more importance regarding the quality of the rubric is time.

Additionally, design is a direct aspect of the quality of the rubric, and it affects both construct validity and reliability. Finally, all these factors function interdependently, nevertheless, the main factor that affects the overall quality of the rubrics are the teachers.

5.1 Research question 1: Which is the formation and knowledge that teachers of English as a second language have at the moment of developing and evaluating rubrics?

First and foremost, we have to address the difference between formation and capacities. The formation that a teacher had is something measurable and certain. Nonetheless, formation and actual capacities are not the same as formation experience does not necessarily reflect teachers' capacities. Still, evidence of the study suggests that formation regarding every aspect of a rubric makes teachers more confident about their own capabilities and decisions regarding the design and application of a rubric. In other words, the greater the knowledge about rubrics, the more experience in the use and application of them, which thus leads teachers to have more developed expert judgment.

Regarding our first research question, we can establish through the answers of the interview that T1 did not have any formation in the design and application of rubrics. In the case of T2 and T3, it was observed that they have training in rubrics but with the limitation of being only about the application of them instead of the whole process that includes designing, applying and giving feedback. In both cases (T2 and T3), they had previous training in calibration of application of rubrics, from two different institutes. As a conclusion, there is no evidence of formation regarding theory of evaluation or rubrics in general for any of the participants. For this reason, we believe it is more suitable to address their "formation" in terms of the experiences they had with rubrics, in terms of design or application, which seemed to have an effect in their expert judgment.

It is important to understand that such expert judgment, (see Section 2.5.1) is largely an *experiential judgment* (as explained by Harris & Eggen, 1993). This is conditioned by the previous experiences that they had. The idea here is that, when a teacher has more experience and training regarding use, design and/or application of a rubric, they make better decisions due to the quality of the rubrics that they made from scratch or the ones that they adapt for a certain kind of activity. In both of the cases examined in this study, the schools provide teachers with rubrics of their own that they can modify.

Firstly, rubrics by T2 and T3 were much more developed than rubrics by T1 in the design and feedback procedures. This was reflected in the quality of the rubrics they used. For example, T2 and T3 rubrics had more components and steps than T1 rubrics; in the rubric “Talking Point” (see Appendix F) that T1 used, there was no presence of levels, and the descriptors were not accurate, whereas the rubric “Fabulous Food” used by T2 and T3 is very accurate in terms of levels and descriptors to evaluate.

Secondly, evidence showed that T1 did less adaptations in the rubrics given to her, whereas T2 and T3 adapted the rubrics more extensively to their necessities, which reflected more confidence regarding their expert judgment.

In the end, the different capacities of the participants seemed to affect the quality of their decision making regarding rubrics use and application. An example of this is the decisions made in relation to the sourcing of rubrics for classroom adaptation. This decision affects the quality of the rubrics and is based on their expert judgment/previous formation. T1 here reported that she uses internet as a source. However, internet is not a source, but a medium, though the webpage in which she gathered the information about rubrics is a source. Conversely, T2 and T3 make use of formal and verifiable sources, which allow the final rubric to be trustworthy. Thus, we could say that T1 is making bad decisions in designing rubrics, such as lack of accuracy of descriptors, due to the absence of formation/expert judgment, and this could result into affecting the quality of the final rubric.

Despite the fact that T1 has less training and experience with rubrics, she has intuitions of what is important when giving feedback which could be translated into validity and reliability. For example, she seems to be aware of the fact that giving general feedback is not an ideal practice and that individual feedback would be. This means that, at least to an extent, she is able to recognize that this practice gives validity and reliability to the evaluation. Nevertheless, as she explains that her time is insufficient, we can see that she gives general feedback to optimize her time.

Furthermore, it could be seen that the descriptors she added in one of the rubrics that T1 modified (see Appendix F) were the most accurate in the rubric. This suggests that there is some expert judgment she has developed through her experience as a teacher that is manifested in her intuition regarding what works best or not.

Regarding piloting, T1 does not show awareness about the importance of the piloting process, whereas T2 and T3 do know how important this is for the process of evaluation. The absence or presence of piloting in the process of evaluation is explained through the previous knowledge and formation both T2 and T3 had about rubrics and the lack of knowledge from T1's side. Since T1 does not have any prior formation regarding rubrics, it seems impossible for her to know the importance of piloting her own rubrics. In other words, her lack of knowledge restricts T1 from knowing about piloting. As for T2 and T3, their previous knowledge about rubrics allows them to be more conscious of what piloting is and how important it is for students to know the task before they have to take the actual test. This could affect the overall quality of the assessment process, and the construct validity of the rubric, however it does not affect reliability.

In conclusion, we can say that teacher's formation in rubrics and their experience is important when it comes to making decisions about them. Evidence in the study points at differences of experience that explain coherently differences in decision-making. In both cases there seems to be an expert judgment in operation that seems to depend on quality experiences and knowledge about assessment and rubrics.

5.2 Research question 2: What are the institutional resources that the teachers can use for the design and application of the rubrics?

In terms of institutional resources the teachers use for the design and application of the rubrics, results showed that two main factors affected the way in which the teachers came about their rubrics: infrastructure and time. We also found that not all the institutional resources that the teachers were provided by the school were equally important for the development of the rubric.

Regarding infrastructure, results showed that there was a basis for the creation of a rubric. This basis consisted in having access to a computer with internet connection. Other resources, such as the type of board the classroom had, or the number of rooms, did not seem to be as important. Therefore, we believe that this did not have an impact in the quality of the rubric as T1, T2 and T3 had all the necessary equipment(s) to create a rubric. Although, this only considers to these two establishments, as they are the ones we are investigating. Thus, it can be said that there is a base ground for the making and application of a rubric regarding the infrastructure of the establishment.

As it was already stated, time seemed to be a critical factor in the designing process and application of the rubric. It seems to us that time could affect the rubric in the sense that T1 only has one hour for planning, and that results in her having to work at home, due to lack of time. This could conceivably affect the quality of the rubric in the sense that she might not have time to search for rubrics, or time to adapt and modify them. Conversely, T2 and T3 have 10 hours for planning a week. Due to this, it seems like they have the time to improve and reflect on their rubrics. There is evidence that, in the case of T1, time was so reduced that it limited the opportunities to improve the rubric. Therefore, this could affect the quality of the rubric indirectly since we can argue that time constraints seem to limit the probability that teachers can improve their rubrics, though this is not a part of the rubric itself but it just depends of the context of the teacher. Nevertheless, even if the teachers had enough time to improve their rubrics, if there is a lack of knowledge about rubrics because of their formation, they would not be able to improve the rubric significantly. Hence, it is not just one factor that affects the quality of the rubric, but an interaction between them (see the reflection in depth in Section 5.4).

We also observed that not all resources are equally important for the quality of a rubric. In particular, it seems that time is more important than infrastructure when it comes to the design and modification of rubrics. Moreover, we have seen that the formation/capacity of teachers is more important than resources as they directly affect the quality of the rubric. As a conclusion, we can gather that the design and use of rubrics do not require many material resources, and only a base ground is needed, which consists in having access to a computer with internet connection.

We also need to acknowledge that there is neither a minimum not a maximum of resources to build up a rubric regarding infrastructure. This is not the case when we deal with time, as we consider that there should be a minimum, proportional to the number of hours they work per week. Regarding infrastructure, it was found that the teachers only needed where to write down the rubric and students to give feedback to them.

We could also find that human resources seemed to be essential, since they reduced time and allowed for feedback in the design and application process. Furthermore, sharing the work made on the rubric and having supervisors checking the rubric was also found to be relevant. In T1's case, who worked on her own, she had less time to distribute among her students in terms of the total number of students she had, the number of classes she conducted, and the planification hours per student she had available. By working together, T2 and T3 could optimize an already reduced time resource. Moreover, that T1 worked by herself may explain in part her lack of autonomy when dealing with rubrics, since she did not have a partner with whom to check or compare her professional work. Under these conditions, the chances of improving the quality of her rubrics (and her teaching, in general) are lower than in the cases of T2 and T3. This is even more so when in T2 and T3's case there is an English department working alongside the teachers, giving them their own feedback and criticizing the rubrics when necessary, which provides further support to the teachers' decisions.

The number of students also proved to affect the feedback the teachers gave. We observed that T1 (who has approximately 40 students per class) mentioned that she stopped delivering individual feedback and changed to delivering group feedback, due to having a big amount of students. Conversely, T2 and T3 deliver individual feedback as they can do it because of the amount of students they have (approximately 20 per class). This, in turn, also affects the evaluating process in its entirety, since in the case of T1, she had to change the way in which she gave feedback because she had not enough time to do it individually. Since each of her classes consisted of 40 to 45 students each, the time she would have to spent giving individual feedback to each of her students was too much in relation to her available time resources. She was aware, however, that giving this type of feedback was an ideal practice, which showed she was able to recognize that this practice gives validity and reliability to the evaluation. Since her time resource was insufficient, she was capable of changing her strategy to still give feedback under her time constraints.

Something that also came up in both cases was the presence of external mandates, namely rubrics that the establishments provided the teachers. Due to being an external mandate, these rubrics were not designed to necessarily match the context in which they are being applied. Furthermore, we found that even though this type of resource was found in both establishments, the quality of the rubrics the teachers were given was different. In T1's case, the rubric appeared to be of poorer quality than the rubric given by the institution of T2 and T3.

Finally, although a baseline regarding resources was observed, we cannot distinguish for certain the rest of baselines, if any. The observed conditions in this study may have an effect in other non-observed factors within this study, for instance we are not considering the external aspect of validity nor surface validity. As we are only focusing on the design and application of rubrics, we only take into account the core aspects of rubrics. We can recognize therefore that a limitation to this study is that the context cannot be fully observed.

5.3 Research question 3: Which is the way in which teachers design and apply the evaluating rubrics to their students?

The discussion of results for RQ3 are here divided into three topics: designing process, application, and feedback. To begin with, the way in which T1 and T2 and T3 designed their rubrics was different. This difference lies in the fact that the former modifies her rubrics by herself, whereas in the case of T2 and T3 they did this process together. Moreover, T2 and T3 have to present their rubrics to the head of the English department (see section 2.5.1 above).

Their rubrics also differed in the way in which they used external sources as a reference to elaborate and modify their rubrics. Even if schematically, their processes are very similar in that they build the rubrics based on external sources. T1 only argues that she uses internet as a source. However, we understand that the internet is not a source, but the medium by which one is able to find sources online. On the other hand, T2 and T3 make use of formal and verifiable sources, such as Cambridge, which allow the final rubric to be trustworthy. This information leads us to believe that T1 does not have any formal sources due to the fact she does not possess any knowledge on rubrics (as reported in in section 4.3 above), therefore, leading her into making inappropriate decisions regarding this process. Thus, we could say that the quality of the sources might, indirectly, affect the overall quality of the final rubric, in the sense that she could be influenced by unreliable sources that she obtains from the internet, and that has an impact on the rubric she designs. Thus, we believe that the sources she used were probably irrelevant.

Regarding the rubrics that were adapted by the teachers, in the case of T1, we can see that in the rubric Talking Point (see Section 4.5.1) the quality of the rubrics seems affected by the fact that there are sub-dimensions that are not properly distributed in the rubric. Therefore, according to our meta-rubric, this could affect the construct validity in a theoretical sense, thus, it has an impact on the quality of the rubric. In addition to this, there are also some aspects in which we can see that the organization of the information is not actually related to content, although it is aligned to the task (see Appendix G). Regarding reliability, as the descriptors are unclear, for instance in the quantity of new words they should use, this would be left to the teachers' criterion, hence, affecting the rubric's reliability as a scoring tool. Therefore, we could conclude that the rubric is poor in terms of its quality.

In the case of T2 and T3, we could see that the rubric that we can relate to T1's rubric as both evaluate oral performances, has a good quality based on our analysis with the meta-rubric. Firstly, because in the rubric (see Appendix F) most of the levels present an exhaustive description. This adds to the theoretical aspect of construct validity and to the scoring aspect, which is related to reliability. Secondly, the descriptors describe clearly what is expected from students to get a certain score, which ensures the reliability of the tool.

Another significant difference between T1 and T2 and T3 is that the latter two adapt their rubrics according to what they observe in the classroom, they contextualize the rubrics, and then they are able to draw different conclusions from what they could observe. Since she does not seem to know about the importance of this practice, T1 only takes into account what she thinks will be better to add or modify in her rubric and her decision has no theoretical foundation. Based on our meta-rubric there is an aspect related to the construct that has to do with the theoretical alignment of the content of the rubrics with the course's program. Therefore, T1's decision, would diminish the quality of the rubric in terms of the pedagogical aspect of the meta-rubric, which is related to construct. Nevertheless, T1's rubrics are actually aligned with the task, which contributes to task-related validity. Conversely, T2 and T3 are certainly aware of the importance of contextualization when dealing with rubrics, whereas T1 is quite unaware of this, hence her rubrics are of lesser quality as they are not contextualized. This is one of the reasons of why we believe the differences they have in their expert judgment might affect the quality of the decisions they take regarding the design and the application of rubrics.

Lastly, T1 evaluates classwork within the same rubric that evaluates the general results of the test, whereas T2 and T3 evaluate processes and results in different rubrics. This finding can also be explained by the lack of knowledge T1 possesses regarding rubrics and evaluation and it is consistent with the formation both T2 and T3 have received (as pointed in Section 4.1 above).

In terms of application, all teachers hand in a copy of the rubric to their students before the evaluation and explain said rubric to them. In addition, T2 and T3 hand in another copy at the same time they give the test as well as uploading the rubric to the platform of the school (as explained in Section 4.3 above). This ensures that the students will have thoroughly studied the rubric before the evaluation as well as having it as they are taking the test, making it a sign of validity. From this we infer that what matters is that the rubric must be delivered beforehand. This does not reflect on any difference of validity/reliability in our meta-rubric. Therefore, we conclude that this does not have an impact in the quality of the rubric. Hence, the difference between the two schools is not as relevant for the resulting quality of the rubric.

As for the way in which the teachers pilot their rubrics, it is said that they do not do it, but in T2 and T3's case, the teachers practice the evaluation before the test without telling their students they are actually preparing for the test (as explained in Section 4.3 above). Based on our meta-rubric the presence of piloting could assure the construct validity of the rubrics since it could help to verify that the rubric is measuring the construct it should assess, besides confirming that it is well designed and that students understand it. Therefore, it helps to ensure that the rubric is well developed. .

Finally, in terms of feedback, we found that time and resources (i.e., the number of students per class) affected the quality of the feedback given. Specifically, time limited its quality. In terms of the assessment process and based on our meta-rubric, feedback contributes to construct validity. A bad description of the levels (see Appendix J) will affect the delivered feedback. In addition to this -according to our meta-rubric- a poor explanation of the scoring process will have an impact on the reliability of the application of the rubric.

Due to the differences in the way T1, T2 and T3 give their feedback, there might be consequences regarding the overall construct validity of the practice, mainly in terms of the consequential aspect of validity. These consequences might affect the fairness of the feedback T1 gives to her student, since she is not able to give a more personalized feedback to each of her students and instead she gives a more general feedback, there might be students whose doubts will not be cleared in their entirety. Therefore, students will have less of a chance to improve for the next evaluation.

T2 and T3 also differ from T1 in that they give a more balanced feedback, both positive and negative, whereas T1 only gives negative feedback. The evidence on which of these two practices is best is limited due to the fact that there is no easy way of knowing if feedback is effective (Margaret Price et al, 2010). Still, evidence indicates that choices of feedback are different between cases. Feedback effectiveness is related to cognitive processes of the way in which students perceive and analyze the given feedback. This means that, if they do not perceive feedback correctly, this will not be effective. Lack of data from the students' perspective in the study does not allow to draw conclusions on this subject.

In summary, in both cases the teachers designed and applied the rubrics in a different way. On the one hand T1 works alone. This is relevant as it makes expert judgment and previous experience prevail in terms of the design of the rubric. On the other hand, T2 and T3 work together. Therefore, they calibrate themselves together. This, in turn, prevents errors regarding alignment with the program. Moreover, they have more time to design and to prevent design errors as they are two teachers instead of just one. The adaptation process was important for this question as we addressed the factor of expert judgment in-depth. In addition to this, we also acknowledged differences -when analyzing the rubric in relation to our meta-rubric- in terms of reliability and construct validity between T1 and T2 & T3. This seems to be affected by the fact that there are more people working in one case. Regarding feedback, we noticed that it has an impact in the quality of the rubric in terms of both the consequential aspect of validity (which is part of construct validity) and reliability. Furthermore, it also affects the quality of the assessment process overall. Finally, and based in our meta-rubric, we were able to analyze that in T1's case alignment with the program is quite low, whereas in T2 and T3's case this alignment is certainly higher and explicit. This reflects either a high or low quality of the rubric regarding the theoretical aspect of the construct.

5.4 Research question 4: What is the relation that exists between the quality of the rubric and the interaction between the formation and knowledge of the teachers and the institutional resources in relation to construct validity and reliability?

The study carried out assumed the existence of a relationship that works based on the formation of the teacher and the institutional resources they dispose of. We found that both of these factors were interdependent. An example of this would be the fact that the resource of time is needed to improve the rubrics, but without appropriate formation on the part of the teachers, the effects of time resources on their own is limited (see Section 5.3)

When it comes to the design of rubrics the factors that affect this process the most are the formation of teachers and their expert judgment. Secondly, in terms of the application of the rubric, the importance that is given to piloting depends on the formation of teachers. Furthermore, the process of delivery of the rubric, related to resources, does not seem to be affected by the resources as long as it is delivered beforehand. Hence, the fact that T2 and T3 have more platforms available to hand in the rubrics is not relevant to the quality of the process. Regarding feedback, the factors that seem to affect it the most are both resources: time and the amount of students. These are the factors that are relevant as they limit the type of feedback to be delivered, and this directly affects the quality of it.

People's autonomy has also a fundamental role in the process of designing and applying a rubric. This would explain the conservative changes made by T1 on the rubric she receives from the institution, since she is limited in terms of exercising her autonomy due to of her lack of knowledge on rubrics as well as being limited by her context. In the case of T1, her context does not allow her to feel more confident in the decisions she makes. This can be interpreted as partly a result of not having colleagues or supervisors (which are part of the human context, see Section 2.5), unlike T2 and T3, who work together in the taking of decisions regarding the final rubric they use. Moreover, by not having any supervisors, T1 does not receive feedback or criticism about her rubrics. By not having colleagues, she is not able either to examine critically her own evaluations in the sense of these lacking something or failing to truly evaluate the task that she is giving her students. In the case of T2 and T3, as they explained, by working together, they are able to calibrate each other's work and support each other's decisions regarding the rubric, thus, they are able to recognize whenever their rubrics are insufficient or in need of changing any aspect.

Hence, we conclude that there is an interaction between the formation of teachers and the human context. Moreover, the human context is related to resources in the sense that if the establishment has more resources it can certainly afford to have more teachers or supervisors to ensure the quality of the rubric, and this could indirectly affect the quality of the rubric in terms of construct validity and reliability.

It was noticed in many cases that formation was more important and had a bigger impact in the quality of the rubric than resources. The latter can limit or restrict certain processes like design, application or feedback (see Section 5.2), however the former will more likely affect the quality of the rubric. One example of this is that not having a person to work with, such as T1 (see Section 4.3 above), opens the way to more difficulties regarding calibration, which in turn can influence directly the quality of the rubric in terms of both design and reliability.

Nevertheless, evidence also shows that working by oneself is still less important than not having knowledge about rubrics. In a hypothetical situation, if T1 had a colleague who works with her in the designing and application of rubrics, she would at least have more insight on the way in which she is designing and applying her rubric. However, if this hypothetical colleague presents the same poor degree of knowledge T1 has of rubrics, i.e. none, having a partner would still not contribute to the overall quality of the final rubric they apply. All in all, the formation that the teachers possess affects the quality of the rubric in that T1 has a less informed expert judgment that T2 and T3, since both of these teachers had had previous formation on rubrics and evaluation.

We can thus argue that it is their expert judgment which leads them to make certain decisions regarding the design of their rubric, decisions which can be more or less informed depending on their formation and practice. In addition, working with a colleague helps in the designing process of a rubric, but if teachers show little capability and have no formation on rubrics or evaluation, it will not matter if such externally-mandated rubrics are of higher quality (as was the situation observed in section 5.2), since every decision they make in their adaptation could have a positive or negative impact on the quality of the final rubric.

During the creation of a rubric, different teachers can follow the same procedures in terms of relying on external sources to elaborate their rubrics, which was the case of T1, T2, and T3. However, if teachers' knowledge and practice about evaluation and rubrics is not adequate and/or insufficient, they could access invalid or unreliable sources (as observed in section 5.1). This may, in turn, end up affecting the validity and reliability of the rubric, and therefore its quality. In the case of T2 and T3, the decisions they make are consistent with their previous knowledge and formation on rubrics and evaluation.

In relation to the design process of rubrics, in the case of T1, classwork is evaluated within the same rubric in which the results of the test are evaluated. Conversely, T2 and T3 evaluate processes and results in different rubrics (see Sections 4.5 and 4.6 respectively). This shows that T2 and T3 have a more thorough knowledge about rubrics and, therefore, have a better expert judgment than T1. In this respect, T1 shows no notion of the way in which different rubrics work in terms of evaluating either processes or results. Since she has no previous formation regarding rubrics and evaluation, T1 has no way of knowing that she should separate these two aspects. In contrast, since T2 and T3 do have knowledge of rubrics and evaluation, they are able to recognize the importance of separating processes from results. This reflects the fact that T2 and T3 are capable to make decisions that reflect a better expert judgment than T1.

Thus, we can draw from this that the quality of the rubrics depends on both the formation of the people who make use of them as well as the resources they have at their disposal. Nonetheless, as we have seen so far, formation seems to have a bigger impact on the resultant quality of the rubric, as every decision teachers make are associated with their knowledge and formation. In contrast, resources may limit or restrict the process affecting the quality of the rubric. T2 and T3 adapt their rubrics according to the specific needs that they observe from their classrooms, they contextualize them, and draw answers from that. This also shows how they are better prepared in terms of their formation, since they observe the learning necessities of each of their students. This, in turn, shows that they are taking better decisions as they consider the context of every course at the moment of designing and adapting the rubrics.

As for resources, since the rubrics are provided to the three teachers as institutional resources, it can be argued that in a context where there is autonomy to manipulate rubrics, the capability and formation of the person will again influence that degree of autonomy that the teacher possesses. In the cases examined in this study, formation seems to give teachers more confidence to make their own decisions which, in turn, seems to impact positively on the quality of the rubric.

However, there were some cases in which we recognized that resources seem to be more relevant teachers' formation as they may completely restrict the process of design and feedback. In particular, it was observed that the quality of the rubrics that were provided by each establishment is unequal. In the case of T1, a poorer quality rubric is provided to her in comparison to the ones T2 and T3 are given, which are supported by established assessment institutions such as Cambridge and the Italian Association (see Section 4.3).

It seems, therefore, that the quality of the starting rubric can also affect the quality of the final rubric the teachers end up designing and applying. This means that, notwithstanding the autonomy and capabilities of the teachers, the quality of the initial resource in the case of an external mandate observed in the cases of this study, affects the overall validity and/or reliability. In sum, the quality of the initial resource seems to limit the quality of the final rubric, independently from the capacities of teachers. This effect would be direct in terms of the way in which it affects the rubric, since if the teacher is given a poorer-quality rubric, the ending result will very likely lack quality.

The quantity of students as well as the time the teachers have in terms of planning hours also proved to be a factor that affected the quality of the rubric as well as the quality of the feedback the teachers provided their students. The former is most important at the time of giving feedback, since due to the quantity of her students, T1 was not able to give a more personalized and individual feedback per student, as was the case of T2 and T3 (see section 5.2). The latter, in turn, affected the rubric in terms of limiting T1 due to her lack of time to make a more detailed work modifying designing her rubrics and allowing T2 and T3 to revise her rubrics plenty of times before they made sure they were completed (see Section 5.2). Thus, in this case it can be seen that despite the formation the teachers may have, they would not be able to implement a better-quality feedback procedure if they do not have enough time to do it.

Additionally, we can observe in this case that there are direct and indirect effects regarding how the overall quality of the rubric seems to be affected. In this sense a direct effect is the one that affects the design of the rubric whereas the indirect effect affects other parts of the process or the person involved in the design (such as time, amount of students per teacher, quality of the received rubric, among others). In contrast to the direct effect, an indirect effect on the rubric's quality was observed in the way in which time, the quantity of students per class, and the received rubric affected the quality of rubrics. Time also depended on other aspects, such as the formation and capacities of the teachers and resources such as the planification hours each teacher was provided with.

Finally, we were able to observe that time resources are more exploitable when there is more than one person working on the same rubric at once. In one case, we have a teacher working by herself (T1), and in the other we have two teachers working in conjunction (T2 and T3). As it was already discussed in RQ2 (see Section 5.2), T2 and T3 working together allows fluency in the process of designing and application of the rubric. Because of this, they advance more quickly in terms of work dynamics, which in turn also means a better validity in terms of construct and alignment with the program and task and in better reliability of both the rubric and the feedback that is provided to the students. We found that being alone and working alongside a colleague did have an effect on the way in which time was to be used by the teachers. T1 working by herself seemed to cause more difficulties regarding time allocation than working with a partner, as was the case of T2 and

T3 (see Section 5.2). T2 and T3 work together and were able to optimize their time in terms of both the design of their rubrics and the way in which they applied them.

Furthermore, we found that time to design rubrics was linked to the formation of the teachers as well, in that the quantity of hours would not matter if the teacher showed no knowledge on the design of rubrics. Specifically, we found that the one hour that T1 had to work on her rubrics was enough due to her lack of formation. If she had had the same ten hours as T2 and T3 had, she would have still the obstacle of not really knowing what to change from the rubric.

After discussing both factors, formation of teachers and resources, involved in the quality of the rubric, we believe that neither the formation of teachers regarding rubrics nor the resources given by the institutions can be studied independently. The evidence in this study suggests that, when the relationship between these two components behaved in a way that they are directly proportional (i.e. there is a high degree of formation and the school provides adequate resources), the quality of the rubric design and application processes seemed to improve. Conversely, when the same two factors behaved inversely proportional (i.e. there is one of the components that is of better quality than the other than the other), the quality of the rubric design and application processes decreased.

Based on this knowledge about how the variables of resources and formation behave, we can conclude that the formation of the teachers is, to some extent, more important than the resources given by the schools especially in the process of design and application of the rubric. In all cases, their previous formation seemed to be directly related to the decisions related to the aspects, such as knowledge of contents and skills, that are going to be present and evaluated in the rubric. This means, finally, that teachers are critical agents in determining the quality of the rubrics' designing process.

To illustrate this point, in the case of T1**,** if she had the same competences as T2 and T3 have, she could have created a better-developed rubric than the ones she currently does, even if she had less time to do so. On the other hand, if T2 and T3 had the same competences as T1 does, even if they had more time to make decisions, the rubric would unlikely be improved. This shows that the quality of the rubric can be attributed mainly to the capabilities of teachers who design and apply them, as observed in their decisions regarding the definition of dimensions and levels, elaboration of descriptors, scoring procedures, feedback practices, etc.

In conclusion, teachers and their capabilities and formation do affect the quality of the rubrics' designing process. If the teachers are better-prepared, they can arguably be capable of designing and developing adequate rubrics. Conversely, it the teacher is not well-prepared in the area of assessment and/or rubric design/application, the decisions they can make will not be well-informed and thus less likely to be advantageous to the final rubric they create or modify.

In the next chapter, we will present the conclusions of our investigation in regards of our four research questions as well as our objectives (see Chapter 3).

Chapter 6: Conclusions

6.1 Introduction

In this chapter, the final conclusion of the study will be presented in different sections. First, a summary of the results is going to be explained based on our four research questions. Then, the limitations of the study will be described, in terms of which variables were or were not considered, so participants and future investigators can take them into account when carrying out eventual explorations on this matter. Suggestions for further research are then offered in order to aid future investigations in the field. Lastly, final comments are made so as to show how this study contributes not only to the field of evaluation, but also to a larger discussion about the enormous inequalities of the Chilean educational system.

6. 2 Summary of results

Results of the study will be presented according to each research question that was formulated. In terms of the formation and knowledge that teacher as a second language have at the moment of developing and evaluating rubrics, two positions are shown. In the first place, T1 reports no previous knowledge about it, only basic knowledge in one class at University was claimed. Overall, formation of teachers seems to be a critical factor affecting the quality of rubrics.

Concerning the institutional resources that teachers can use for the design and application of rubrics, T1 was provided with only a basic teachers' room and Chromebook laptops. On the other hand, T2 and T3 were provided by computers, laptops, projectors reported several technological resources (laptops, projectors, etc.) as well as an exclusive office for the department of English, and free access to library and photocopier, including a platform to share material among students and other teachers. Results suggest that time is the main resource that affects the quality of rubrics.

In terms of the way in which teachers design and apply the evaluating rubrics to their students, T1 was provided some by the school itself, while others were taken from the internet by the teacher and adapted according to what she considered were the necessities of her students at the moment. T2 and T3 worked together in the application of rubrics. They created and adapted them, in order to later present them to the department, and only approved rubrics were able to be applied. They do not use the same rubrics every year, as they are constantly updated. The ones that are adapted were taken from a reliable source. Results indicate that there are substantial differences in terms of construct validity and reliability in the two institutions observed. The rubrics used by better-prepared teachers in the more resourceful institution were clearly more valid and reliable than the rubrics from the institution that has less resources and less-prepared teachers.

Finally, in terms of the relation that exists between the quality of the rubric and the interaction between the formation and knowledge of the teachers and the institutional resources in relation to construct validity and reliability**,** the observed interactions worked interdependently. In some cases, formation affected quality regardless of resources. Other resources affected quality regardless of the formation of teachers. Finally, formation resulted in being critical for the rubrics' quality, and time showed to be the most important resource as it affected most interactions between formation and resources.

6.3 Limitations of the study

This study has some limitations, as certain variables were not analyzed due to shortcomings of time and resources. Moreover, they were not considered because the specific aims of the investigation did not deem them to be absolutely necessary. All the explanations hereunder will clarify what these limitations are and further describe why they are in fact a limitation.

To begin with, there are some factors to bear in mind to avoid considering this study as a representation of all Chilean schools, and to steer clear of generalizations. First of all, the investigation is a case study, which means that the two specific cases that were taken into account were thoroughly reviewed and studied, and in no case, should they be generalized, since we are only studying those two particular cases and their context. The sample is unrepresentative of the variety of schools in the country. The spectrum is not wide enough to let us know the reality of the rest of the schools in Santiago, let alone in Chile. Besides, since it was a contextualized study, it does not necessarily reveal what can be expected of a public and a private school. Finally, the number of interviewed actors is too low to be representative and so the study does not necessarily reveal the way all teachers are trained concerning the creation of rubrics and their application in schools.

Second, it must be considered that the gathered information is incomplete in relation to the contexts we described. This has mostly to do with variables that surround the assessment process such as other actors, the teachers' formation, institutional resources, context, and the actual use of rubrics.

Regarding other stakeholders relevant to an assessment situation, only the teachers' self-report was considered, but there are many more actors involved. These are the students, the school's management, students' parents or legal guardians, and other stakeholders. This implies a limitation because other important relations that go beyond the teachers and their formation in relation to resources could not be analyzed (for example, the role of the Ministry of Education and other education-related institutions). There is a much larger spectrum of information surrounding the process of design and use of a rubric, which can only be contributed to by other actors.

Students, for instance, are directly affected by the use of rubrics, since the grades they get can determine their future academic decisions, namely, if they continue their education after school and, if so, where. With this in mind, it would be useful to know whether they understand the rubrics and whether they deem them fair or even useful, among many other possible questions.

Furthermore, it is the students who receive the feedback. Therefore, they might have a more accurate idea about its real effectiveness or, at least, a different perception that could contribute to the discussion. The fact that all our information is drawn from the interviews with the teachers, and that they are the only stakeholders we interviewed, means that there is information missing from other stakeholders to enrich and validate the evidence found in this thesis.

In the same line, regarding the teachers' formation, it must be noted that we only have the information they gave us about their studies regarding rubrics. This information is biased, since all we know about the evaluation courses the teachers attended is their own perception about it. We did not access to the teachers' CVs nor their certification documents or diplomas and so we are not informed about their performance, their attendance, and their grades in evaluation courses. Therefore, we have no way of evaluating these courses' quality.

Regarding institutional resources, we only considered the information given by the teachers regarding the number of their classes, students per class, levels, and paid planning hours. This limits the study, as we can only speculate whether their case is the same as other English teachers in the institution that teach other classes. However, we do not have access to other sources to know for certain by interviewing only one or two of them. In order to turn this around, interviews to other English teachers of the school is to be expected, as well as other stakeholders of the school, such as principals or directors. Moreover, we would have to gather information about all of the rest of the teachers' classes, since their workload also influences the way in which they come about their evaluations and rubrics in specific courses and levels.

Regarding context, we are not aware of the socioeconomic context of the students. Again, we can only speculate. We previously established that rubrics are, in a way, the description of a task. For the evaluation to be as fair for everyone as possible, then, the design of such task should be adapted to the context in which it takes place. Our analysis considers context, as it aims to be critic of Chile's economic inequity, so patently reflected in the differences in the quality of instruction through different sectors of the country. However, interviewing two schools with seemingly different backgrounds is not enough to generalize by claiming that the students' context is clear. More needs to be considered in order to confirm a more solid relation between a high socioeconomic context and a better quality of the rubrics, and a low socioeconomic context and a poorer quality. To analyze context in depth, other pieces of information are relevant, such as the students' living situation, including information such as the parents' income and whether the student is a parent him or herself. The quality of the rubric depends also on the effect it has in the learning of students, which, in the Chilean context, depends heavily on the conditions in which students live.

Finally, regarding the actual use of rubrics, we did not make direct observation of lessons. This is a limitation, since actual lessons are genuine practices of the use of rubrics. We only saw what the teachers gave us: the actual rubrics they worked with, and their personal statements of how they applied the rubrics. Also, the only rubrics analyzed were oral, and used in second year of high school. It would be ideal to carry out an in-depth study which is aimed at investigating all of high school's levels in order to better understand the way in which the rubrics work depending on each different level. Moreover, being able to analyze rubrics of written evaluations would allow us to make a comparison between oral and written rubrics and study their similarities as well as differences.

All these limitations imply there is a wider spectrum that is worth observing when investigating the functioning and the shortcomings of evaluation and rubrics in Chilean institutions. If those limitations were considered and analyzed, more aspects regarding the use of rubrics and their application could be understood in depth. It would also provide a more complete picture of the situation of Chilean education and its differences across different socioeconomic contexts.

Nonetheless, the gathered information is certainly significant, as it allowed us to establish rather clear relations. The limitations did not mean an impediment, as our aim was not to observe the implementation of rubrics itself, but to understand the relation between teachers' formation, institutional resources, and the quality of rubrics in *segundo medio*. Furthermore, although we were not able to register some data, we can still assume that the information provided by the teachers is trustworthy enough for the type of relations we sought to establish.

6.4 Suggestions and further research

With enough resources, the described limitations can be overcome in a second or third version of the study. Further research might include interviewing the other actors, accessing to the other actors' information, registering lessons and feedback occasions, and expanding the study to written evaluations. Moreover, we suggest piloting and interviewing twice, and asking follow-up questions about the rubrics and the course programs. Further ways to overcome the limitations might consist on investigating other institutions and a wider spectrum of contexts, observing the use the teachers give to the planning hour(s), accessing to the rubrics the teachers download from the internet or accessing the web pages themselves, conducting a focus group between English teachers or students from different institutions, and handing in a written questionnaire or poll to the teachers or to other actors.

6.5 Final comments

This study has two fundamental characteristics. First, we chose to conduct research beyond the reality of the University and the English program, and aimed to investigate a context where political and educational inequalities are more in display. This decision was also based on the fact that the corpus of assessment studies about universities is much more extensive that the ones about schools and, more specifically, high school. Second, we were interested in observing differences across real schools and to truly grasp this reality.

This study contributes to evidencing how institutional resources, the formation of teachers, and the quality of the rubrics are not only interrelated, but they are also interdependent with one another. Thus, we demonstrated that there is a group of reciprocal relations that are established into three variables. This implies that the characteristics of one of these variables will always affect the remaining two.

Furthermore, this investigation contributes to understanding the studies regarding evaluation beyond theory. Through case studies, we are able to take into account factors that do not necessarily have to do with the design of rubrics, but that are a part of the reality of each establishment and, therefore, are essential when evaluating the quality aspect of rubrics. To ignore these factors means to study evaluation through an idealistic or unrealistic lens, where an ideal school, teacher, and students are part of the scenario. In line with this, there should not be any interruption or external factor to consider when designing a good-quality rubric, and it would be enough to study the elements that constitute the rubric as such, like validity and reliability.

Since this investigation includes context and its importance, it contributes to further comprehending the study of rubrics as an inevitably social matter. The analysis, moreover, contributes to seeing evaluation as a political matter, since resources and formation are directly linked to the educational and economic policies of each country, as we have seen through this study.

In this sense, the investigation reported in this thesis is instrumental in evidencing how an unequal context affects the quality of the rubrics. Clearly, the study contributes to a larger discussion about the inequities of the educational system, which is fundamental and critical in light of the current social unrest in the country and the strong questioning of the neoliberal experiment that has been conducted in Chile since the military dictatorship.

The implementation of this type of economy brought colossal inequalities to the educational system, which became private. These neoliberal principles have resulted in a stratification that has become one of the major concerns of the public opinion. Indeed, good-quality education is only accessible for Chileans who enjoy a favorable economical status, which is a rather reduced group of people. This thesis thus contributes to the fundamental discussion of how the lack of resources results in the lack opportunities for teaching for institutions that cannot provide resources for good quality education.

REFERENCES

AERA, APA, & NCME. (1999). Standards for educational and psychological testing. Washington DC: American Educational Research Association.

Bachman, L. F. (1990). Fundamental Considerations in Language Testing. Oxford: Oxford University Press. Ch. 7 Validation, pp. 236-295

Brown, H. D. (2004). Language assessment. Principles and Classroom Practices. London: Pearson/ Longman. Ch 2: pp. 19-41

Chavez Eras, A., & Florez Guzman, M. (2017). Desigualdad Educativa y el Neoliberalismo en la Educación Chilena. *INNOVA Research Journal*, 2(3), pp. 167-174.

Cornejo Chávez, R. (2006). El experimento educativo chileno 20 años después: una mirada crítica a los logros y falencias del sistema escolar.

English, C., tests, E., & Advanced, C. (2019). C1 Advanced exam format | Cambridge English. Retrieved 24 December 2019, from https://www.cambridgeenglish.org/exams-and-tests/advanced/exam-format/

Harris, D., & Eggen, P. (1993). The impact of experience on conceptions of expertise: A comparison of the thinking of veteran, first-year, and preservice teachers. In *Annual Meeting of the American Educational Research Association*, Atlanta.

Luoma, S. (2004). *Assessing speaking*. Ernst Klett Sprachen. pp. 103-105

Measurement Error. (2006). Retrieved from https://socialresearchmethods.net/kb/measerr.php

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), pp.741-749.

Messick, S. (1996). Validity of Performance Assessment. In: G. Phillips, ed., *Technical Issues In Large-Scale Performance*, 1st ed. Washington DC: National Center For Education Statistics, p.7.

Moreno-Doña, A., & Jiménez, R. G. (2014). Dictadura Chilena y Sistema Escolar:“a otros dieron de verdad esa cosa llamada 1 educación”. *Educar em revista*, (51), pp. 51-66.

Moskal, B., & Leydens, J. (2000). Scoring Rubric Development: Validity and Reliability. Practical Assessment, Research & Evaluation, 7(10), pp. 4-5.

Muñoz, D. (2009) Reliability as a Context- Dependent Requirement for Writing Proficiency Assessment. Language Studies Working Papers, 1.

Pérez, E., & Martínez, C. Validez de contenidos y juicios de expertos: una aproximación a su utilización. Rev Avances en medición.[Internet]. 2008 [consultado 1/09/2017] 6: 25-34.

Price, M., Handley, K., Millar, J., & O'donovan, B. (2010). Feedback: all that effort, but what is the effect?. *Assessment & Evaluation in Higher Education*, *35*(3), pp. 277-289.

Souza, A. C. D., Alexandre, N. M. C., & Guirardello, E. D. B. (2017). Propriedades psicométricas na avaliação de instrumentos: avaliação da confiabilidade e da validade. *Epidemiologia e Serviços de Saúde*, *26*, pp. 649-659.

Stevens, D. and Levi, A. (2005). *Introduction to rubrics*. Sterling, Va.: Stylus. pp. 3-93

Trochim, W. 2006. Social Research Methods, retrieved 3 September 2019, from https://socialresearchmethods.net/kb/truescor.php

Trochim, W. 2006. Social Research Methods, retrieved 3 September 2019, from https://socialresearchmethods.net/kb/reltypes.php

True Score Theory. (2006). Retrieved from https://socialresearchmethods.net/kb/truescor.php

APPENDIX

Appendix A: Interview

1. ¿Cuántos cursos tiene a cargo? ¿Cuántos niveles por curso?
2. ¿Cuánto del tiempo de la planificación usa para el diseño y aplicación de la rúbrica?
3. ¿El colegio le entrega posibilidades de perfeccionamiento profesional o investigación respecto al diseño y aplicación de rúbricas? ¿Cuáles?
4. ¿Cuál es la infraestructura y equipamiento que provee el colegio para el diseño y aplicación de rúbricas?
5. ¿Qué de la infraestructura y del equipamiento que tiene disponible usa para el diseño y aplicación de la rúbrica? *(pregunta complementaria)*
6. **¿**Tiene horas de planificación pagadas por el colegio?
7. Si la respuesta es no. ¿Cuántas horas de su tiempo personal invierte en la planificación? (*pregunta complementaria)*
8. Si la respuesta es sí. ¿Cuántas?
9. ¿Tiene algún conocimiento previo sobre el diseño y aplicación de rúbricas?
10. Si la respuesta es sí. ¿Cuáles y de dónde las obtuvo? *(pregunta complementaria)*
11. La rúbrica, ¿la hace usted mismo o la obtiene de otro lugar?
12. Si la obtiene de otro lugar, ¿de dónde?
13. ¿Es la rúbrica adaptada conforme a las tareas/evaluaciones?
14. ¿Ocupa una o más rúbricas?
15. ¿Quiénes participan en la elaboración de la rúbrica?
16. ¿Qué actividades se realizan en la elaboración de la rúbrica?
17. ¿Cuál es la utilidad que le dan a la rúbrica?
18. ¿Se realizan procesos estadísticos para garantizar validez?
19. ¿Se alinea la rúbrica a la tarea que esta describe?
20. ¿Cuándo se lleva a cabo este proceso?
21. ¿Cómo se lleva a cabo?
22. ¿Pilotea sus evaluaciones para garantizar validez?
23. Si es así, ¿Quiénes la pilotean? ¿Cómo?
24. ¿Cuáles son las prácticas respecto al feedback que otorga? Ejemplo: subrayar, marcar errores, dar feedback oral, etc.
25. Dentro de las prácticas mencionadas, ¿cuáles consideraría más efectiva según su experiencia?
26. En evaluaciones escritas, ¿cuáles son las áreas en las que se enfoca al dar el feedback? Ejemplo: gramática, puntuación, contenido, organización, etc.

Appendix B: Piloting questions added to the interview

1. ¿Hubo preguntas que no quedaron claras? ¿Cuáles? ¿Por qué?
2. En términos generales, ¿Cuál es su opinión sobre la entrevista?

3. En cuanto a la duración, ¿Considera que la entrevista fue extensa?
4. Desde su apreciación, ¿Agregaría o quitaría alguna pregunta de la entrevista?
5. ¿Considera que la entrevista estuvo bien conducida por el entrevistador?
6. ¿Introdujo el administrador el tema sobre el cual se basaría la entrevista?
7. ¿Tiene usted alguna sugerencia o comentario para esta entrevista?

Appendix C: Piloting Interviws' transliteration

E: Ya (pausa). En primera instancia, ¿Cómo estás?

I: Bien. Un poco agotada, estresada por la evaluación docente, pero en general bien.

E: Perfecto.

I: Sí.

E: Bueno, como te contaba, esta entrevista consta de preguntas que van a ir en base a saber un poco más del proceso de evaluación que se hace acá en el Liceo. Eh, y esta entrevista es grabada obviamente. Así que, mm, ¿te sientes preparada para la entrevista que vamos a hacer?

I: Sí, lista.

E: Lista y dispuesta.

I: (risas)

E: Perfecto. ¿Cuántos cursos tienes tú a cargo y cuántos niveles por curso?

I: Son seis cursos, tres niveles.

E: Ya.

I: Dos cursos por nivel.

E: ¿Y cuáles son esos niveles?

I: Primero, tercero y cuarto medio.

E: Perfecto (pausa). ¿Cuál, eh, cuánto es el tiempo de la planificación que usas para el diseño y aplicación de una rúbrica?

I: Eh, desde que empiezo a hacerla a terminarla unos tres días yo creo.

E: Perfecto. Para el diseño y la aplicación de la rúbrica.

I: Ah, claro. La aplicación viene después.

E: Perfecto. Eh, este cole- ¿el Liceo te entrega posibilidades de perfeccionamiento profesional o investigación respecto al diseño y aplicación de las rúbricas? Y si es así, ¿cuáles?

I: Eh, nos han intentado ofrecer desde la municipalidad, pero por las situaciones de todo más la pérdida de clase, etcétera, no hemos podido aplicar ninguno.
E: Ya, perfecto. Em, Respecto a la infraestructura cuáles, ¿cómo es la infraestructura y el equipamiento que provee el colegio para el diseño y la aplicación de las rúbricas?

I: Eh, bueno como cuando hacemos rúbricas generalmente son para presentaciones orales o trabajos, eh, nuestros mayores recursos son tener data en la sala, parlantes en todas las salas y el recurso humano que son las curriculistas que son quienes nos corrigen las rúbricas cuando las mandamos.

E: ¿Y las hacen acá dentro del Liceo?

I: Sí, acá adentro.

E: Ya, perfecto. ¿En la sala de profes?

I: En- sí. O en el departamento de inglés.

E: Ya. ¿Tienen esas dos posibilidades?

I: Sí.

E: Perfecto. Em, (pausa) ya. ¿Tienes horas de planificación que el colegio te pague?

I: Sí.

E: Ya, ¿cuántas?

I: Las horas no lectivas correspondientes que son 35% si no me equivoco en la carga oral.

E: Y mm, ¿tienes algún conocimiento de previo sobre el diseño y aplicación de las rúbricas?

I: ¿Como desde la Universidad?

E: O cursos, o cosas que tú hayas tomado.

I: No, solamente, eh, ocupábamos diseños que buscábamos en libros o fuentes y a partir de eso cuando las enviábamos las las curriculistas nos iban corrigiendo las cosas que estaban mal y ahí podía aprender a hacer rúbricas.

E: Pero ningún ramo así como específico para la evaluación.

I: No.

E: Ya (pausa). Ya, respecto a la rúbrica en sí, ¿la haces tú o la sacas de algún otro lugar?

I: No, las hago yo
E: Perfecto. Em, ¿y estas rúbricas que tú haces las adaptas al tipo de evaluación, o es una para todas las evaluaciones?

I: No, depende de la evaluación porque hay algunas que son orales, otras escritas y orales, otras solo escritas.

E: Ah, perfecto. Ocupan rúbricas para las presentaciones orales y las escritas.

I: Sí, claro.

E: Maravilloso. Y ocupai, claro, dos tipos de rúbrica, o más tipos de rúbrica no solo una.

I: Sí, o la misma rúbrica incluye distintas habilidades que tienen que mostrar en un solo trabajo E: Perfecto. ¿Y quiénes participan en la elaboración de las rúbricas?

I: Eh, generalmente es trabajo colaborativo, como con las mismas colegas de inglés, em, nos aportamos cosas y después las mandamos para que las revisen en UTP.

(03:43) E: Y siempre las revisan las curriculistas.

I: Siempre.

E: Perfecto. Em, ¿qué actividades se realizan en la elaboración de la rúbrica?

I: En qué sentido?

E: Eh, se reúnen primero, se reúnen todas juntas.

I: Ah, claro. Empezamos como decidiendo qué tipo de trabajo vamos a hacer. Qué habilidades vamos a evaluar, etcétera, y después empezamos a hacer como una checklist de las cosas que queremos evaluar. Eh, y a partir de eso ir buscando los criterios que vamos a…

E: Perfecto. Em, ¿se realizan procesos estadísticos para garantizar la validez de la rúbrica que ustedes usan?

I: No.

E: Ya. Y esta rúbrica que ustedes hacen se alinea con la tarea que ustedes quieren evaluar?

I: Claro, el objetivo que queremos de las niñas.

E: Perfe. Eh, ¿cuándo se lleva a cabo este proceso de rúbricas? Una vez al mes.

I: La confección?

E: Claro.

I: Eh, bueno antes era muy poco como que lo hiciéramos, hace tres años, por ejemplo. Eh, este año, eh, mucho. Hemos tratado de diversificar la evaluación de los cursos, no hacer solo pruebas escritas, así que hemos hecho muchas., mm, si de cinco notas por ejemplo que pusiéramos en el semestre, dos o tres son aplicando una rúbrica.

E: Perfe (pausa). ¿Pilotea sus evaluaciones para garantizar validez? Es decir, por ejemplo, si vas a tomar una prueba oral, antes de tomar la prueba tú hací ejercicios orales con las niñas?

I: Claro, sí. Todo lo que se evalúa se practica antes en clases.

E: Perfe. Y eres, es la profesora encargada del curso quien hace este ejercicio previo a la evaluación.

(05:28) I: Sí, en clases.

E: ¿Cuáles son las prácticas respecto al feedback que se otorga? Subrayar, marcar errores, das feedback oral, etcétera.

I: Em, bueno lo típico ejercicios de guía o cuaderno, eso se revisan uno por uno, las niñas se acercan y se corrigen y se explica el error que tenían. Y en el caso de las evaluaciones orales, em, ocupamos mucho como board games y claro las tengo que escuchar, ir puesto por puesto y escucharlas e ir corrigiendo los errores.

E: Bacán. Y dentro de las prácticas que tú mencionaste, ¿cuáles consideras más efectivas según tu experiencia?

I: Em, en cuanto al aprendizaje, las rúbricas si son buenas porque,bueno es complicado porque con la rúbrica como es muy guiada, eh, es mucho más fácil de realizar, pero hoy día las niñas están mucho más al aplicar algo, como dibujar, presentar algo que completar una prueba por ejemplo. Entonces hoy día las actividades prácticas manuales son mucho más efectivas que hacer una prueba de memoria por ejemplo.

E: Entonces el feedback que podí dar como oral respecto a esas evaluaciones te sirve más.

I: Sí, sirve más.

E: Perfe. Y en las evaluaciones escritas, ¿cuáles son las áreas que se enfocan en dar feedback? Gramática, pronunciación, o sea, puntuación, contenido, organización.

I: Eh, mayormente gramática y use of English, eso.

E: Ya, perfe. Ahora te voy a hacer preguntas respecto a la entrevista, a las preguntas que te hice anteriormente.

I: Listo.

E: Eh, ¿hubo preguntas que no quedaron claras? ¿Cuáles y por qué?

I: Eh, hubo una, que creo que te pregunté la explicación que eh…

E: El proceso o no?

I: Sí, creo, el proceso.

E: Ya, perfe. Y en términos generales ¿cuál es tu opinión sobre la entrevista, sobre lo que yo te acabo de preguntar.

I: Em, me llama la atención que se enfoque tanto a la rúbrica porque es algo que se ocupa mucho, que nosotros estamos utilizando mucho ahora en el Liceo.

E: Perfe. En cuanto a la duración, ¿Considera que la entrevista fue muy extensa?

I: No, no, para nada.

E: Ya, eh, Desde tu apreciación, ¿Agregaría o quitaría alguna pregunta de la entrevista?

I: Mmm, no, no, creo que no.

E: ¿Considera que la entrevista estuvo bien conducida por mi (el entrevistador)?

I: Si (risas) Si.

E: Eh, ¿Introdujo el administrador el tema sobre el cual se basaría la entrevista?

I: Si, también.

E: Y, ¿Tiene tu alguna sugerencia o comentario para la entrevista?

I: Eh, Quizás solamente hablar del resultado de la aplicación de una rúbrica.

E: Ya.

I: Claro, como partir de la experiencia eh, qué tan positivo o negativo es utilizar rúbrica en cursos poblados como los del liceo.

E: Claro si, por que igual hay una diferencia entre un curso de 45 alumnas y un curso un poco más pequeño.
I: Claro, de 20 alumnas por ejemplo, el trabajo es distinto a, en cuanto al mismo feedback que uno tiene que darle, es mucho más extenso en grupos chicos, que tener un curso de 45 con 10 grupos por ejemplo de 4 alumnas, eh, es más, toma mucho más tiempo evaluar por ejemplo estas cosas toma dos o tres clases, entonces eso, sobre todo a nosotros que perdemos clases, eh lo hacemos pero igual eh no tanto, como quisiéramos, porque como estamos contrarreloj siempre, y esto toma trabajo en clase, que pueden ser dos clases y además la evaluación son tres clases más, ya pierdo prácticamente un mes, es una sola evaluación.

E: Claro, en aplicarla, en crear la rúbrica, aplicarla y el feedback.

I: Exacto.

E: Y crees que, porque hay muchos colegios donde no usan las rúbricas, o donde no es predominante ese uso como evaluación, ¿tu crees que igual las rúbricas hacen que tus alumnas se sientan mucho más tranquilas de que a todas las van a evaluar de la misma forma?

I: Claro, y se sienten tranquilas porque está muy claro que es lo que se evalúa en el trabajo, en la evaluación que tengan que hacer, porque es explícita, entonces es como una lista de cosas que yo tengo que cumplir, por eso les va mucho mejor también, porque todo está muy específico en una rúbrica.

E: Entonces ha mejorado como el rendimiento. (al unísono)

I: El rendimiento, sí.

E: Y ¿tus alumnas conocen la rúbrica antes de la evaluación?

I: Siempre.

E: Ya.

I: Siempre se presenta la rúbrica, se lee punto por punto y para resolver dudas, y ellas se quedan con la rúbrica en mano, entonces en la, durante la aplicación del trabajo, surgen más dudas y van con la rúbrica aclarando.

E: Perfe, listo estamos listas muchas gracias.

Appendix D: Interviws' transliteration Scuola Italiana

T3: Entonces nosotras como que tomamos un molde y de, ah... hay algunas evaluaciones que nos gustan, por ejemplo no se "un video de no sé qué" y esa… la rúbrica la hemos ido... oye entonces la Sonia me dice que nos faltó ver esta parte...

T2: La adaptamos

T3: Que se nos pasó la arreglamos pal proximo año y así las hemos ido <ininteligible> modificando.

T2: En el fondo ves las necesi… o sea, las características del curso también y a base de eso tú vas modificando o adaptando las rúbricas de los cursos. Ahora claro en la universidad te enseñan a hacer rúbricas pero uno a medida que to.. los conocimientos que uno adquiere ponte tú… yo hace más de quince años salí de la universidad entonces uno tiene que ir, las evaluaciones que se hacían en esa época no son las mismas que se hacen ahora. Entonces también tienes que ir adaptando porque cuando yo salí de la universidad no era muy común hacer videos por ejemplo. O cuando mis colegas más antiguas, nuestras colegas más antiguas salieron de la universidad, hacer un video era una cosa asi como imposible.

T3: Te adaptas, sí.

T2: Y ahora cualquiera puede hacer un video, entonces tu tienes que ir adaptando e ir aprendiendo cosas nuevas y muchas veces todos los conocimientos están en internet.

T3: Lo que es investigación también. Cuando tú pides investigar ya no se pide investigar a una biblioteca, es poco común. Aparte que lo que buscamos nosotros no está en la biblioteca, generalmente está en línea la… la información. Y vamos adaptando todo al, a lo que queremos, al objetivo, a un objetivo que tenemos. Qué es lo que vamos a hacer.

T2: Nosotros básicamente acá nuestras rubricas hasta segundo medio son las de... estas son las de FCE. Porque como tenemos que llegar a cierto nivel del… como el European Framework, entonces tenemos, necesitamos tener rúbricas de eso. Y se ocupan también en primero se adaptan, por ejemplo aquí, no se po, aquí, esta es la del FCE, sí? Entonces por ejemplo cuando dice…

E1: ¿Qué es la FCE?

T3: First English Certificate, es el de… B2?

T2: B2.

T3: B2 del Common European.

E1: Ya.

T3: Cuando ustedes salen de la universidad ahora deberían salir con CAE, que es lo que tengo yo ahora.

E2: A mi me lo dieron en el colegio.

T2: Claro.

T3: ¿En serio?, Oay, y yo lo saqué el año pasado.

<risas>

T3: Ya, con ehm, con eso, ya? Porque ahora los colegios están pidiendo de CAE para arriba, entonces...

E2: Qué brígido.

T3: Los enanos acá deben salir con un… FCE, sí? Entonces, por ejemplo… pero uno lo va adaptando al nivel, por ejemplo aquí dice no se poh, eehm... “communicate straightforward, complex ideas as appropriate” entonces, por ejemplo, no se po, un niño de primero medio no le vai a pedir que haga un análisis como de un… así como profundo tanto una idea, sino que… lo haga lo mejor que puede y… eh como usando estructuras complejas en relación ah… a su nivel.

E2: A su nivel.

T3: Ya, entonces ese lo… este lo ocupamos en primero y segundo, si?

T2: Segundo más que nada, en primero esta se baja.(05:27) E1: Estas son sobre los recursos del colegio, eh, y su relación con los recursos del colegio. La primera es ¿Cuántos cursos tienes a cargo?

T2: Tú empieza.

T3: Yo, Paola. Tengo seis.

E1: Ya, y ¿Cuántos niveles separados son por curso? O sea, como de niveles distintos.

T3: Cuatro niveles. Ah no, tres niveles. Tengo dos segundos, dos terceros y dos cuartos.

E1: Ya.

T2: Yo también tengo seis cursos, tengo dos quintos básicos, un primero, un segundo, un tercero y un cuarto medio.

E1: Ya.

E2: Qué variado.
E1: Sí. Eh, ¿Cuánto del tiempo de la planificación usa para el diseño y aplicación de rúbricas?

T3: Ya, Paola. Yo tengo 10 horas de planificación pero comúnmente para las rúbricas usaremos eh, tres horas, dos horas <voces externas>. Ya, de las 10 horas de planificación yo más o menos ocupo dos o tres. Las que tenemos los días miércoles porque son las horas que tenemos como seguidas, juntas y comúnmente yo como el miércoles no tengo clase, tengo solo permanencia, entonces me dedico los otros días a corregir pruebas y cosas.

T2: Ya sí, Sonia, eh, al crear una rúbrica o al hacer cualquier instrumento de evaluación también las horas que les damos, no podemos decir, "ya, me voy a sentar y voy a terminar esto", porque puedes, le dedicas ponte tú dos horas a algo, lo dejas, guardas el material y después vas al día siguiente, lo revisas y lo sigues trabajando. Le puedes dar una hora más, le puedes dar la tercera hora, yo creo que le podrías dar hasta una cuarta hora de revisión al material. Tanto solas como juntas porque acá los materiales, no sé si te lo habíamos comentado, los materiales, las pruebas, las rúbricas, las hacemos en conjunto, porque es para el nivel, entonces lo ideal es que todo el nivel sea evaluado de la misma forma, igual que las pruebas, se hace el esqueleto de la prueba y se hace para los dos cursos lo más parecido posible. No importa... lo más similar, que es lo más justo para todos, atendiendo a las diferentes características de cada curso igual.

E1: Ya.

T2: Y de cada alumno.

E1: Ya. ¿El colegio le entrega posibilidades de perfeccionamiento profesional o investigación respecto al diseño y aplicación de rúbricas?

<I1 y I2 al unísono> No <risas>

T3: Es que no. O sea, el año pasado como que quisieron hacer algo pero al final trajeron a una señora que hablaba de las rúbricas y mostraba ejemplos de cómo ella había aplicado la rúbrica y todo, pero no fue un taller en donde yo me siento y digo, "ya, a ver, yo necesito evaluar tal y tal cosa, háganme una rúbrica y vamos corrigiendo", sino que fue una señora hablando de lo maravillosa que eran sus rúbricas.

E2: Ya.

T2: En ese caso es mejor bajarlo de internet y traerlo a la realidad.

E2: Sí. <risas>

(08:37) E1: Ya, ¿cuál es la infraestructura y equipamiento que provee el colegio para el diseño y aplicación de rúbricas?

E2: Como las salas.

T3: Computador con internet.

T2: Sí, computadores con internet. Bueno las salas, para poder evaluar a los alumnos, tenemos las salas con data, con parlantes con buen audio porque hay buena amplificación en la sala. Entonces si los alumnos hacen un video nos permite evaluarlo bien, no es una cosa que hay que traer parlantes para poder escuchar mejor el video, no. Eh, ahora ahí los alumnos no lo pueden echar la culpa a que el data está malo o que el computador está malo o que la amplificación está mala y que por eso su video tuviera una mala nota. Ahí hay algo bueno para poder evaluarlos y que se ajuste.

E1: Ya, perfecto. ¿Qué de la infraestructura y del equipamiento que tienen disponible usa para el diseño y aplicación de la rúbrica? Como que realmente lo usa, no que solo lo tienen.

T3: Los computadores y el data. Es que claro el data... o computador porque para lo escrito no necesita nada más. Nuestra sala de profes y sería.

T2: Sí.

E2: Que es un espacio, igual tienen departamento de inglés.

T3: Claro, tenemos departamento de inglés.

T2: Tenemos a disposición...

T3: La biblioteca.

T2: La biblioteca, fotocopias también. Hay acceso a fotocopiar las, a que tú le hagas la rúbrica de tu computador al alumno, uno la puede hacer personalmente ahí por alumno y después mandarla a imprimir, o mandas a imprimir todas las hojas en blanco y después vas rellenando las rúbricas. No hay...

T3: Ah sí porque nosotros les entregamos después del trabajo a cada uno su rúbrica con su puntaje.

E2: Ya.

T2: Porque antes se les dice, se evaluarán de esta forma, se les muestra a ellos a través con el data uno les va mostrando, “ya, les vamos a evaluar esto, tales cosas, y se evaluarán de esta forma, aquí está el puntaje”, todo. Se transparenta. Se puede subir a la plataforma también.
T3: Ah sí, eso es importante.
T2: Porque hay una plataforma School Track y ahí los alumnos saben que está ahí la forma de evaluación, la rúbrica. Y después la podemos imprimir y se les entrega en papel, o sea tienen acceso a ella tanto electrónico como físico.
E2: Tienen de todo. <risas>
E1: Ya y, ¿tienen horas pagadas de planificación por el colegio?
T3: Sí, las horas dependen de tus horas de clases, se saca un porcentaje...
T2: Sí, hay una proporción.
T3: El 40%, entonces por ejemplo yo tengo 24... 29 horas, entre cinco de reunión y 24 de clase. Y de eso se saca un porcentaje que es el 40% y te dan tus horas de permanencia. Entonces en ese caso yo tengo 10, porque tengo esa horas de...
T2: Yo no sé cuánto tengo, pero si tienes 10 horas de clase te corresponden cuatro horas de permanencia.
(11:45) E1: Ya, perfecto. Ya, después estas son respecto a las experiencias previas sobre rúbricas y cosas así. Eh, ¿Tiene algún conocimiento previo sobre el diseño y aplicación de rúbricas?
T3: Sí, por la U y por el Británico, porque en el Británico a pesar de que nos daban hechas las cosas como que igual nos hacían retrabajarlas, mirarlas, usarlas entre... como para, no entre nosotros pero así como… “y qué corregirías en este”. No, de hecho nos mostraban por ejemplo alumnos que estaban dando FCE y nos pasaban la rúbrica. Entonces tú veías el video y después ellos te decían así como “ya, qué puntaje tú le darías, tanto, por qué le diste tanto, qué errores fueron los que tomaste en cuenta para eso”, entonces yo por lo menos de eso-
T2: Sí, yo también, eh, aprendí de la universidad, lo que se usaba en esa época, que se adapta ahora al presente, y también trabajé en instituto tomando exámenes PET. Entonces para eso te hacen capacitación con la rúbrica de ellos, y como decía ella hacen lo mismo, con videos Cambridge te manda una vez al año se hace acá en Chile una capacitación de cómo evaluar a un alumno, y evaluarlo con los estándares de ellos. Entonces te pasan la rúbrica, te pasan el estándar de ellos, tú ves el video y lo mismo que decía ella, ves el video y le tienes que dar la nota, y después ellos descubren, te muestran la nota que dio Cambridge y hay que ir comparando, y ahí te van entrenando en cómo aplicar la rúbrica, las razones por las cuales el alumno debería tener esta nota, qué tan cerca estás tú, y vas comparando y te van dando un feedback inmediatamente sobre lo que tú deberías hacer o lo que Cambridge espera del desempeño del alumno.
E2: ¿Y en cuanto a diseño?
T2: En cuanto a diseño no, las envían hechas y dicen...
T3: Porque es la misma para todos, es la misma para todo el mundo. Cambridge manda la rúbrica que ocupan ellos en todo el mundo para los exámenes internacionales. Entonces, por ejemplo la que tenemos acá que ocupa el examinador cuando tú das el FCE.
E2: Y ustedes la adaptan igual a las necesidades de...

T3: O sea, esta de acá está tal cual. Lo que nosotros hacemos cuando evaluamos, hacemos una adaptación. No sé, por ejemplo no esperamos que el niño en segundo medio, no se po, use passive voice, perfect tenses, sino que una estructura compleja de acuerdo al nivel de ellos.

(15:04) T3: Nosotros, como dice la Sonia, como estamos en el colegio, tenemos la opción de moverlo.
T2: Flexibilizar un poco.
T3: Exacto.

E2: ¿Igual tienen como sus criterios para esa flexibilización?
T3: Sí.
E2: Ya, o sea, conversan entre ustedes y todo
T3: Sí, es que están por ejemplo dados acá po, acá mismo, por ejemplo...
E1: Después vienen esas preguntas.
E2: Ah perdón
T3: ¿Ah?
E2: Me adelanté
T3: Entonces, por ejemplo, acá cuando dice “ideas complejas”, eh después de ver las unidades uno sabe qué tipo de ideas complejas son. Yo creo que ya, es como, lo tení como integrado. Como que no tenemos que hablarlo, así como “ya, si menciona esta idea compleja…”, no, es como que ya nos fluye.
T2: Ahora tú ya, también va en parte del conocimiento del alumno. Porque uno los conoce. Cambridge por ejemplo cuando toma un examen, es: pasa una persona, next, salió, next, next, y es como una fábrica <risas> es como una pequeña fábrica donde tú vas evaluando no más, evalúas y chao, evalúan, chao.
E1: Claro.
T2: Aquí también influye el conocimiento del alumno.
E2: Como personal.
T2: Personal. Acá influye la evaluación diferenciada que tienen muchas veces, entonces ahí uno tiene que guiarse, “ah, tengo la evaluación diferenciada”, entonces aquí hay ciertas cosas que no le puedo tomar a un alumno, o alguna dificultad que tenga, ponte tú, un alumno que sea tartamudo.
E1: Claro.
T2: Eh, ahí uno no puede aplicar en la pronunciación el estándar que te da, sino que, claro, lo conozco, pero el alumno está, sí, expresando una idea, sí está expresando un pensamiento, hay una reflexión, pero está ese pequeño impedimento que es, eh, la tartamudez, que no se lo consideras en la pronunciación para bajarles nota.
E2: Claro, claro.
E1: Ya, entonces sería...
T2: Aplicas criterio, en el fondo.
E1: Como, de la universidad y del Británico, o de Cambridge, como de instituciones aparte que también…
T2: Claro.
E1: Ya.

T3: ¿Y en el IB no te enseñaron…?
T2: En el IB, claro, cuando uno, es que trabajé en un colegio IB, también, y en los colegios IB...
E1: ¿Qué es IB?
I1 y E2 <al unísono>: Bachillerato Internacional.
E1: Ah ya.
E2: ¿En cuál?
T2: Eh, británico de Punta Arenas.
E2: Ah, no era el mío <risas>
T2: No, no era el tuyo <risas>. Y también, en el IB se trabaja en base a proyectos, entonces de ahí tienes que elaborar bien a, em, eh, capacitarte en cómo evaluar a los alumnos, y también se evalúa mucho en base al desempeño del alumno y eso lo evalúas con rúbricas.
E1: Ya.
T2: Entonces, ahí tienes que ir creando con tu paralela de acuerdo a la unidad que están haciendo, porque también la unidad, eh, tiene que nacer de los alumnos, de la indagación que quiera hacer el alumno, eh, va, va naciendo y tú tienes que ir elaborando como... Tú tienes una evaluación general, como una rúbrica general y esa la tienes que adaptar a la unidad que vas a tomar, que vas a tomarle a tus alumnos, y evaluar el proceso, porque en ese caso se utiliza rúbrica para evaluar el proceso de aprendizaje. Lo evalúas completo, no haces una prueba, una prueba de la unidad de “caballos”, y también se integra con las otras asignaturas.
(19:46) E1: Ya, y el último aspecto es sobre las rúbricas mismas, primero, eh ¿La rúbrica la hace usted misma o usted mismo o la obtienen de otro lugar?
T3: Miti mota, o sea, parte la sacamos de algún sitio de internet donde había algo más o menos cercano a lo que queríamos nosotras evaluar. Porque la idea es común, o sea cuando nosotras planeamos como la evaluación de la unidad es como, ya, “qué es lo que queremos, queremos que hagan esto y que usen esto, esto y esto de tal forma”, ¿sí? Entonces después nosotras buscamos así como en Internet, a veces la Sonia me dice “mira, encontré esta” y yo digo “oh encontré esta otra” y vamos pegando, juntando, y la vamos haciendo nuestra, al final.
E1: Claro
T3: Como un Frankenstein de muchas cosas.
E1: ¿Y usted?
T2: Lo mismo, sí.
T3: Trabajamos juntas <risas>
E2: El colegio no les da rúbricas entonces.
T2: No.
E2: Ya.
T3: Ah, hay unas que nosotras nos...<risas>
T2: ¿La busco? ¿La trajimos?
T3: No, sí la tengo <risas>.
T2: La trajiste, ya.
T3: Lo que pasa es que, eh bueno, acá en el colegio los niños dan un, un examen de estado, entonces bueno ya no dan la Terza Prova. Entonces ya no tienen que escribir en inglés, pero sí, por ejemplo, eh nosotros seguimos manteniendo la rúbrica de la Terza Prova que era en

ese tiempo, como nuestra, para este tipo de preguntas, preguntas abiertas, ¿sí? Porque es la... hasta ahora ha sido la que más se acerca al desarrollo, pensamiento y uso del conocimiento como argumento, que es lo que queremos lograr con este tipo de preguntas, que va direccionado al examen de estado, entonces, esta la robamos... <risas>
T2: Sí, es un robo.
E2: Está en italiano. <risa>
T3: Y la, y la mantuvimos para nosotros. Sí, está en italiano, no la, no... nunca la hemos querido traducir porque...
T2: Se entiende.
T3: Sí y los alumnos entienden sin problema, no siempre la entienden, pero, ellos, pero es fácil de...
T2: Y muy clara.
T3: Es muy clara.
T2: Es... no es extensa, y es súper clara porque, claro...
E2: Sí, hasta yo entiendo.
T2: Hasta tú entiendes, ¿viste?
T3: Si no es compleja.
T2: No es compleja, y que mide conocimiento, competencia y capacidad.
T3: Y la otra que también nos robamos es la del “colloquio”, que es la del examen oral.
(23:43) T2: Tomamos prestadas.
T3: Tomamos prestado, esta que es la rúbrica que se usó en el colloquio del año pasado.
(27:20) T3: Entonces, rúbricas que nosotras hemos tomado como literales, y usadas así como sin ningún cambio son: las de FCE, que son *speaking* y *writing*, que es la de allá, y esta es la del examen de estado que van con las pruebas de los libros.
T2: De los libros.
T3: Todas las demás están modificadas según la evaluación.
T2: Según las necesidades nuestras. Porque no existen aquí.
T3: ¿Se las puedo mostrar o no?
E1: Claro.
T3: Espérame, la del video, les voy a mostrar la del video.
E1: Entonces toman como de internet.
T2: Sí, ideas.
E1: Algunas referencias, y después la adaptan a lo que ustedes necesitan.
T2: A la necesidad, a lo que nosotras necesitamos.
(29:17) T2: Esta la tomamos de internet.
T3: Esta de internet pero la fuimos moviendo, ya?

(29:40) T3: Por ejemplo, a veces venían así como muy cortas y como que no había una… como una progresión. Entonces por ejemplo la Sonia me decía “oye pero acá mira, en esta parte está mencionado, no se po, ‘sound is not clear’, y aquí no”. Entonces ahí lo agregábamos y lo poníamos como al medio “is fairly clear”, y le fuimos cambiando cosas hasta que quedó… igual hay que agregarle otras.

(30:19) T2: Uno les va dando los tips antes, por eso tienes que presentar la rúbrica antes.

E2: ¿Siempre se las muestran antes?

T2: Sí.

(32:14) T3: Entonces todo eso hay que irlo agregando, como más cosas todos los años, hay que re-mirarla. No es como que hemos podido usar la misma rúbrica todo el año. Estas sí po, porque están precisas.

E1: Claro.

T3: Pero las otras hay que…

E2: Hay que adaptarse.

T3: Ahí hay que revisarlas.

E1: ¿Y estas están hechas por la Scuola o por los colegios de Italia?

T3: ¿Esas? Esas las mandaron, fueron las del colloquio del año pasado, las trajeron de Italia.

(33:35) E1: Ehm, hay algunas preguntas que ya fueron respondidas en verdad, como por ejemplo si las rúbricas se adaptan a las tareas y evaluaciones, sí.

T3: Sí.

T2: Sí po.

E1: Si ocupan más de una, también.

I1 e I2 al unísono: Sí.

E1: Ehm, ¿Quiénes participan en la elaboración de la rúbrica? Como… ¿Ustedes dos?

T3: Por nivel, sí.

T2: Sí.

T3: Es que por niveles se hacen las rúbricas para evaluaciones, entonces como con la Sonia trabajamos segundo, tercero, y cuarto, nosotras hacemos las…

T2: En el fondo siempre creas la rúbrica con la persona con la que compartes el nivel.

T3: El paralelo

E2: Siempre hay un grupo de trabajo.

T2: Siempre, claro. Siempre trabajas con tu paralelo y se presenta al jefe de departamento, para que el jefe de departamento le dé el visto bueno final, estar en conocimiento de. Pero siempre se trabaja con tu paralelo, nunca la rúbrica es la rúbrica de la Paola, o la rúbrica de la Sonia, o la rúbrica de la Marcela.

E1: Ya.

T2: Es la… no es personal, sino que…

E2: Todos de acuerdo.

T3: Sí.

T2: Para la evaluación y están, si estamos de acuerdo como…

T3: Como equipo

E1: Ya. Y se la muestran a la… a la jefa de departamento…

T3: Sí.

E1: Que da el visto bueno y ahí la pueden empezar a usar.

T2: Sí.

E1: Ya. Ehmm… ¿Qué actividades se realizan en la elaboración de la rúbrica? O sea como, eehm, ¿Cómo es el proceso de elaborar la rúbrica? Bueno, básicamente ya lo explicaron también.

T2: Sí.

E1: Que es como… tomar unas que ya están, ir agregándole cosas nuevas, ¿Eso lo hacen en sus horas de planificación?

T3: Sí.

T2: Primero tienes que tener claro qué es lo que vas a evaluar. Si no sabes qué vas a evaluar no…

T3: Es que como llevamos tanto tiempo trabajando juntas es como…"oye y el video sí, ya, mira queremos esto y esto, a ver déjame hacer un borrador, a ver tú también, ehm", lo vemos ahí juntas.

E1: Ya.

T2: O sea uno anota en un cuaderno de lo que, de las cosas que se pueden ir evaluando

T3: No necesitamos, no necesitamos sentarnos, así como por una hora para crear la rúbrica sino que a veces trabajamos juntas.

T2: A veces en el auto por ejemplo, "oye se me ocurrió que podríamos evaluarles tal cosa"

T3: Agreguemos esto y pah.

T2: Claro, sí.

T3: Ejecutivas.

E1: Y, ¿Cuál es la utilidad que les dan a las rúbricas? O sea como, ¿Para qué la usan, para qué, eem, procesos de su evaluación usan la rúbrica?

T2: Se, se utiliza para evaluar el proceso. Porque por ejemplo la del video se utiliza para evaluar el proceso de trabajo en clase, de la investigación, y el resultado final. Se evalúan… y hay otras que, por ejemplo…

T3: Esta también es como de, como también evalúa… No, ésta no evalúa proceso. Lo que pasa es que nosotros en esta, es que en algunas, nosotros pa las que son solo "writing" hacemos una como práctica. Pero esa no está evaluada.

(37:26) T3: Las dos que dicen project tienen como evaluación de proceso y resultado.

T2: Y de resultado.

T3: En las demás resultado, pero el trabajo en clases no se evalúa.

E2: Ya, perfecto.

E1: Ya… ¿Se realizan procesos estadísticos para garantizar validez?

T3: Mmm no, yo nunca he hecho procesos estadísticos.

T2: No… Lo que sí, vemos a qué le podemos dar mayor o menor importancia, eh... más o menos peso.

(38:33) T3: Yo creo que en realidad nunca lo hemos hecho así como de números ni de informes, pero por ejemplo, no sé, por ejemplo, en tal prueba: "Sonia, sabís qué en este ítem está muy difícil o sabís que en este ítem está super bueno y les encantó, ya hagamos, repitamos ese pal próximo año o pa otra evaluación"

E1: Ya

T3: Pero así como estadísticas de números no, no hay tiempo.

E2: Es más difícil.

T2: No y no hay tiempo para hacer estadísticas.

T3: No y en inglés específicamente es complejo porque es muy abstracto… o sea, claro, el video pero el video también, no sé po, ves inglés, diferentes niveles de inglés... Entonces no es como una lengua madre por ejemplo que en español tu sabis que todos hablan español entonces todos tienen que manejar ciertas cosas… En inglés es más complejo.

E1: Ya… Eh,, ¿Se alinean las rúbricas a la tarea que esta describe? También dijeron que sí. Ehm… ¿Cuándo se lleva a cabo este proceso? Cuando están… o sea, como antes de empezar la rúbrica y de recaudar la información se enfocan en la tarea que van a evaluar y a partir de eso hacen la rúbrica.

I1 e I2 <unísono>: Sí.

E1: Ya. Eh, ¿Pilotean sus evaluaciones para garantizar validez?

T3: No. Es que yo creo que eso lo hemos hecho a través del tiempo. Eh, así como saber qué es lo que resulta y qué no, como pa' que sea válido y ver el resultado que esperamos. Eh… no hemos piloteado ninguna, pero sí por ejemplo hay veces que a los niños, que a un curso les cuesta más alguna prueba y hay que como hacer una recuperativa pero... y después ya tú sabes qué tienen que reforzar más el próximo año pa' que no vuelva a pasar, pero como pilotear no… No.

T2: No la piloteamos porque tenemos claro lo que vamos a evaluar y cómo se desempeñan los alumnos, entonces no…

T3: Es que lo otro como decía Sonia... como que los conocemos mucho.

E1: Claro.

T3: Entonces sabemos de lo que son capaces de hacer y hasta dónde los podemos llevar.

E2: Igual son cursos chicos, ¿cierto?

T3: Veinticinco.

T2: Veinticinco, veinte ¡Pero se portan como cincuenta! <risas> Pero uno sabe hasta dónde los podis' presionar pa' que den. Entonces, eh… yo sé de lo que son capaces entonces no, no tengo problema.

E1: O sea como que nunca han hecho una prueba que no vale como prueba para pilotear, pero si usan la experiencia de cada prueba para mejorar las otras.

T3: Sí.

T2: Claro.

E2: Igual tienen esta actividad de Bowie por ejemplo, que durante las clases hacen otro, o no?

T3: Ah, sí, claro <sonidos ininteligibles> Y con esa misma tenemos el "opinion essay".

T2: Sí. También cuando hacen "opinion essay" ellos saben…
T3: Les apunté... Esta... la rúbrica del opinion essay, y esta es la práctica que nosotros hacemos. En el día de la prueba se les entrega la instrucción y líneas, sí? Pero en la práctica ellos tienen que practicar con las partes de la, con la estructura.

E2: Ah ya, perfecto.

T3: Con la estructura, sí? Entonces uno les corrige así como "oye pero esto no tiene nada en relación a esta cuestión, oye pero en la conclusión me escribiste cincuenta palabras y no explicaste qué onda, tenías que hacerlo", etcétera. Entonces al final se les entrega esto y una hoja llena de líneas.

T2: Con el tema.

T3: Sí.

T2: Con el tema a desarrollar.

E1: Entonces ¿Suelen practicar como la tarea antes de la prueba?

T3: Solo el "writing".

E1: Ya.

T3: Los demás no. No, porque por ejemplo, es que, no es, por ejemplo para los orales no es una práctica intencionada. O sea nosotros sí, nosotros sí, pero no les decimos “ya vamos a practicar hoy para el oral exam”. No, es como “oye ya pero dame una idea, pero me estay dando un concepto, te estoy pidiendo una idea, profundiza, dime esto, usa esto otro”.

T2: Y todo eso lo vas haciendo clase a clase también, cuando uno les da un concepto en clase “ya, discutamos esto”.

E2: Algo más natural.

T3: Claro.

T2: Exacto, estamos viendo lo que... si lo estás viendo en literatura uno toma un concepto de la unidad y terminas de ver una... leer un par de páginas y “bueno veamos qué conceptos son importantes acá”, o para empezar también. <ininteligible>

T3: La idea es que ellos lo hagan por ellos mismos y que no sientan que están como estudiando para una prueba, no se si se entiende.

E1: Claro, sí.

T3: Por que si uno les dice “ya, esto”, ellos saben que lo van a practicar para la prueba y después se olvidan, entonces si con las partes orales hiciéramos lo mismo, para ellos dejaría de ser algo como natural de como un progreso que tiene que tener una persona que está en la media, ¿sí? Entonces sería como “ah ya, hoy dia vamos a tener clases de oral exam, entonces la otra clase no hablo”. Entonces la idea es ir sin decirles, nosotras si sabemos obviamente, intencionar la práctica para los exámenes.

E1: Ya.

E2: Ay que son buenas profes, me encantó.

<risas>

E1: Ehm, ya. Las últimas son respecto al feedback. ¿Cuáles son las prácticas respecto al feedback que otorga? Por ejemplo, subrayar, marcar errores, dar feedback oral, etc.

T2: Depende de la evaluación.

T3: Sí, voy a buscar mi cuadernito.

T2: Sí, depende de la evaluación. Tú, por ejemplo, si es una evaluación oral, una interrogación, le das el feedback, al momento que termina la evaluación le das el feedback. Le dices “mira, esto hay que mejorar”, o “lo estás haciendo muy bien, el próximo año va a subir el…”, por ejemplo ahora estamos haciendo ahora justo unas evaluaciones y les estamos diciendo como “ya, el próximo año te toca cuarto medio, entonces ya necesito que desarrolles más tus ideas, te falta un poco ahi”, ehm, “necesito que seas más concreto al momento de dar un… ahm, ¿Cómo se dice?, un fundamento, tienes que fundamentar tu respuesta con lo que estudiamos”, “tienes que comparar cuando te pido comparar”, o a algunos les das un tema y te hablan de otra cosa. Entonces al final, los paras durante la presentación o al final les dices “¿Sabes que? yo te pregunté tres veces sobre el tema y tú insististe en hablar de otra cosa”, pero tu les dices al mismo momento la nota, porque tienen acá, ehm…

E2: Ahí mismo ¿Al tiro pone la nota?

T2: Sí, sí. Después de la interrogación les damos al tiro la nota en general.

T3: ¿De qué?¿Del oral exam?

T2: Sí, del oral exam, porque les damos al tiro el feedback.

T3: Te traje... es que tengo mi cuaderno.

T2: Ella tiene todo ahí. Cuando tenemos videos, le anotamos en la hoja, les ponemos notas sobre lo que… el desempeño del video. Tomamos notas en un cuaderno, y después se traspasan las notas al oral exam, o sea a la rúbrica de evaluación del video, entonces uno le pone “el volumen era muy bajo”, le vas poniendo las observaciones. Cuando es ehm... a ver, ay ¿Qué otro tipo de evaluación tenemos?

E1: Los writings.

T2: Los writings, se van poniendo los símbolos.

T3: Si ah... es que te puedo traer, no tengo la de segundo ahora pero tengo de tercero medio.

E1: ¿Qué cosa?

T2: La hoja con los símbolos, que dice “word order”, los “editing symbols”.

E1: No pero… no es necesario, mientras lo expliquen.

T2: Es que usamos símbolos de edición, para el writing.

T3: Que se los pasamos desde primero medio. Así como “sp” ellos tienen que saber que “spelling” es porque… lo malo que tienen ahí es porque está mal escrito.

E2: Claro.

T2: Claro, subrayamos donde haya un drama en “spelling” y ellos lo tienen que reconocer, si hay un problema. Subrayamos, ponte tú, ya, toda esta línea tiene malo el orden de las oraciones, de las palabras pero… estaría bien, ya, le marcas ahí, le pones un “word order” y le pones “WO”,

E1: Mmm, ya.

T2: Y ellos lo van leyendo y saben ya, y les vamos poniendo notas también.

T3: Sí, a los lados.

T2: No, no se entiende. En algunos casos subrayas y un signo de interrogación: “no se entendió la letra”, y ellos saben que cuando hay un signo de interrogación es que la letra no se entendió.

E1: Ya.

T2: Y también ellos saben que esas cosas van a ir bajando nota.

E2: Ya.

T3: Entonces al final cuando uno entrega la prueba, les da como, uno les da como la base de lo que uno esperaba que contestaran en cada pregunta, y después los llamai de a uno. O sea, por ejemplo, esta prueba que... nosotros ya hacemos una sola al principio porque vienen así como con el cerebro lento, entonces en segundo medio les hacemos una sola prueba de estas de como antiguas.

T2: Materia.

T3: ¿Ya? De materia. Entonces claro, estas partes, es... tenís bueno o tenís malo. O sea, no… no hay de otra.

E1: Claro

T3: Pero por ejemplo en estas, en las de los libros que tienen estas preguntas de desarrollo uno les dice y después los llamas de a uno. Ya, “¡Pepito bla, venga para acá. ¿Tiene preguntas? Sí, Miss, lo que pasa es que no entiendo porque esta pregunta me bajó”, ya.

T2: Si yo contesté bien.

T3: Claro, contesté...

T2: Si le contesté todo.

T3: Entonces...

T2: Me sabía el libro y lo contesté de memoria.

T3: Sí, entonces uno le dice "claro oye, pero aquí me hiciste un resumen, y yo no te estoy pidiendo un resumen, te estoy pidiendo razones de…".

(50:11) l2: Pero sí lo anoté, los errores que más me llaman la atención, porque tampoco es que anoto todo, los que más me preocupan o los que se repiten, y los anoto, y les digo "oye, ya, súper bien, pero ojo este, ojo con este porque manso chamullo", y eso es... lo que se le da a los niños, ese es como el feedback.

(52:01) T3: El "fact file", el video, si desconté algo lo pongo, por ejemplo, este... estos alumnos, a pesar de que tuvieron un video super bueno... ehh, no estaba toda la información que estaba en su hojita de investigación, que debía estar en el video. Y, ehh acá, usaron español en el minuto 01:23, ¿Sí?. Entonces después dicen "¡aay!, pero...!", "ya, revise su video el 01:23". Y acá se los explico "ya, mira, esa es la de 'speaking'", que después bueno ahí lo ven, entonces por ejemplo hay cosas... Yo la mayoría de las veces trato de no ser tan asi como: "oye sabí que tu fluidez es como, es mala" y qué sé yo. No, sino que por ejemplo: "te falta fluidez, en ciertas partes donde te pusiste muy nervioso trata de…".

(53:16) T3: Pero ponerlo de una forma más positiva, sí. Así como: "relájate un poco más… trata de no, no hacerlo, de no ponerte tan nervioso", y ahí ya como que "encourage ", no sé como decirlo en español.

T2: Alienta.

E2: Alentar.

T3: Alentarla a que ¡super bien!, que en vez de ponerte nerviosa o mirarlo, porque de verdad que él es así como "¡waa!", "eh, oye, ya pero cálmate, vamos a hacer esto", de... de acogerlo y que se vió en el video, ¿sí?, eh, eso.

(54:53) E1: Ya, eh, y dentro de las prácticas mencionadas de feedback, ¿Cuáles consideraría más efectivas según su experiencia?

T3: Mmm, a que, como…

E1: Como por ejemplo, entre pasarles una hojita con los puntos o llamarlos uno por uno, o dar una idea general al curso.

T3: Ah no, uno por uno.

T2: Sí, personalizado.

T3: O sea hay... por ejemplo esta, da lo mismo si la hací uno por uno, la alternativa uno por uno... general. Porque eso no es y se acabó, entonces uno les explica por qué es esa y no las otras, pero esa es la misma para todos. Pero la personal yo creo que es la más … como es que, cambia un poco la percepción del alumno de lo que está haciendo. Algunos de hecho te agradecen así como "Oh bien miss, ya, voy a poner más atención a eso la próxima vez" y lo cambian, otros te dicen solo "sí", y siguen haciendo lo mismo, pero al final tú como que les das... les das su espacio. Es como su momento para ellos de, de saber cómo lo están haciendo y como también ayuda la conexión entre el profe-alumno.

T2: Ahora, son... generalmente la respuestas que son también de tono más personal, donde ellos tienen que re-elaborar, entonces, también ellos valoran más que tu le des directamente el feedback, y se tiene que hacer de esa forma porque no puedes hacer un feedback en general sobre una... sobre una… una respuesta tan personal que da el alumno.

E1: Claro, ya... Y, la última es: En evaluaciones escritas, ¿Cuales son las áreas en las que se enfocan al dar el feedback? Por ejemplo: gramática, puntuación, contenido, organización, etc.

T3: Todas.

<risas>

T3: Es que, gramática y...

T2: Es que depende de lo que estás evaluando, entonces estás evaluando un ensayo... el ensayo de opinión, tienes que darle importancia también a la gramática, a la puntuación, a "voice", porque si les dices "ya escribe, escribe un ensayo formal", tienes que utilizar vocabulario formal, y son instrucciones que se dan, no es una cosa que se dé por hecho. Muchas veces la puntuación, en lo que es escritura, no muchas veces, siempre, la puntuación influye en qué escribiste.

E1: Sí.

T2: Si no lo pones puntuación y pones un bloque de palabras, si pones... no utilizas párrafos, el, el sentido de lo que está el alumno tratando de comunicar se pierde. Y el, el... el público, la audiencia

T3: Te mal… mal… "misunderstands".

T2: Claro no entiende, tú tampoco no entiendes lo que está, lo que lees, lo que, lo que escribió el alumno. Entonces sí influye lo que tú... si estamos evaluando una, mmm, una... un texto, hay que tomar eso, hay que… tienes que darle la misma importancia.

E1: Ya.

T3: Sí. Ah pero como en pruebas como en la de los libros, que son estas respuestas como abiertas, la gramática, a pesar de que se evalúa, no es lo más importante.

E2: ¿Lo que importa es el contenido?

T3: La idea, y de hecho la rúbrica, por ejemplo…

T2: Por eso se utiliza esa rúbrica.

T3: Acá, sí. Porque esta rúbrica... la parte del lenguaje va acá, dentro de los seis puntos, pero aparte también va cómo utiliza la idea, entonces en realidad estos seis puntos serán tres... tres de quince los que están relacionados con gramática, qué sé yo. Entonces al final es cómo comunica la idea y cómo utiliza al libro como argumento, eso es lo que me interesa más que cómo lo escribe.

T2: El centro no es la gramática.

T3: No, aquí no, entonces sí tenemos evaluaciones donde sí es porque es necesario evaluarlo, pero esas, esas son las pruebas de “writing” específicamente, y tienen su rúbrica especial para eso.

E1: Ya.

Appendix E: Interviws' transliteration Escuela Arturo Toro Amor

T1: Perfecto.

E: Bueno, esta entrevista eh, consta... es un set de preguntas abiertas donde la idea es que tú puedas contestar con lo que tú pienses etc. Eh, estas preguntas van a estar relacionadas a la rúbrica de evaluación.

T1: Ya.

E: Y lo que nosotros estamos tratando de investigar es cómo afectan eh, el instrumental eh, y el, como ehm... los instrumentos que provee el colegio...

T1: Ya.

E: Con tus propias capacidades como profe de inglés en la calidad de la rúbrica.

T1: Ya.

E: Entonces a eso van a ir enfocadas eh, las preguntas principales.

T1: Mmhh.

E: Entonces la primera pregunta es que cuántos cursos tú tienes cargo y a cuántos niveles por curso.

T1: Ya, yo tengo nueve cursos. Tengo un quinto b de la básica y el resto son media, son ocho. De primero a cuarto medio, y son a y b, entonces serían en total nueve cursos.

E: Ya… perfecto.

T1: ¿La cantidad de alumnos también?

E: Sí <simultáneo>

T1: Sí <simultáneo> bueno emm... aproximadamente cuarenta.

E: Por curso.

T1: Sí.

E: Ya.

T1: Sí, en en quinto básico son cuarenta y cinco.

E: Mmhh.

T1: En primero son menos, treinta y siete por curso, cuarenta.

E: Ya.

E: Igual está como <ininteligible>.

T1: Sí.

E: Ya. ¿Cuánto tiempo de planificación usas tú para el diseño y aplicación de rúbricas?

T1: Eh bueno, en el colegio nos entregan una rúbrica para ciertas cosas.

E: Ya.

T1: ¿Ya? Entonces yo lo que hago es modificarla.

E: Mmhh.

T1: Ya, que me demoraría como... cuánto... ¿Unos veinte minutos?

E: Ya

T1: Porque tengo que ver qué cosas voy a agregar, qué cosas voy a quitar, y en lo demás cuando yo tengo que hacer una rúbrica me demoro más poh'.

E: Ya.

T1: Como aproximadamente... ¿Cuarenta minutos?

E: Ya, perfect.

T1: Sí.

E: Y ¿El colegio te entrega posibilidades de perfeccionamiento profesional o investigación respecto al diseño y aplicación de rúbricas?

T1: ¿De rúbricas así como tal? No.

E: Ya.

T1: Pero siempre hay capacitaciones a fin de año o a comienzo de año.

E: Ya.

T1: Pero es más como lo que ellos te, te ponen.

E: Ya.

T1: ¿Me entiendes? No es como que yo pueda...

E: Elegir.

T1: O sea, se pueden elegir algunas cosas...

E: Mmhh.

T1: Pero no hay tanta variedad tampoco.

E: Ya. Y, respecto y… formación respecto a evaluación, ¿Tampoco?

T1: ¿De evaluación? Ehh... sí, sí, también.

E: Ya.

T1: Pero yo recuerdo que escogí otra cosa.

E: Ya.

T1: Sí.

E: Perfecto. ¿Cuál es la infraestructura y el equipamiento que provee el colegio para el diseño y aplicación de rúbricas? Dígase ¿Ustedes tienen una sala como profesores de inglés? ¿O tienen una sala como pa' los profes? ¿Han tenido un lugar como pa' poder trabajar?

T1: Ya, hay solamente sala de profes.

E: Ya.

T1: Está dividido por la media y la básica. La básica está abajo más grande y la media está en el segundo piso. Eh, sería eso en realidad y nos entregan un computador.

E: Mmhh.

T1: Que es <interrupción: sonido de celular> que es un chromebook, que es uno chiquitito que tú trabajas como con cosas online.

E: Ya.

T1: ¿Ya? Por ejemplo archivos online, todo online, entonces nada se pierde.

E: Ya.

T1: Y con ese computador eh, la verdad es que trabajo casi todas las clases. Lo uso para proyectar, para trabajar en el colegio o a veces en la casa también.

E: Ya, perfecto. Ehm... y esa es la misma, ¿Es el mismo equipamiento que tú tienes disponible para la generación de rúbricas? o ¿Las rúbricas que haces tú las generas en tu propio computador, en tu casa?

T1: No, lo hago ahí mismo.

E: Ya.

T1: Que la verdad es que ya me acostumbré, al comienzo era un poco difícil.

E: Ya.

T1: Porque no es el mismo sistema que un compu normal.

E: Ya.

T1: Pero como ya me acostumbré y lo empecé a hacer así, y como se va guardando, de repente no traigo el compu y lo uso el computador del colegio.

E: Ya.

T1: Entonces entro a mi Gmail y ahí está en Google Drive.

E: Mmhh.

T1: Y ahí lo saco y lo termino de modificar para imprimir y cosas así.

E: Perfecto. Eh... el colegio eh, ¿Tú tienes horas de planificación en el colegio?, y ¿Te las pagan?

T1: Tengo una.

E: Una hora de planificación.

T1: <risas> Sí <entre risas> una hora.

E: Ya, y ¿Te alcanza con esa hora de planificación?

T1: Por supuesto que no. <risas>

E: Ya.

T1: No, no me alcanza. De hecho creo que a todos los profes nos dan una...

E: Ya.

T1: Pero en algunos profesores tienen otras horas de, no sé poh, de gestión, o de jefatura, entonces algunos profes tienen más horas como entre comillas eh, libres que obviamente no son libres.

E: Sí.

T1: Pero sirvan pa' trabajar.

E: Claro

T1: En mi caso yo tengo solo una y trabajo en, en las ventanas que hay, porque como las horas son siempre cronológicas y en las clases son eh, pedagógicas ahí se hace como ese rango de minutos.

E: Tiempo.

T1: Entonces ahí trabajo con las ventanas.

E: Ya, eeh ... <inteligible> <risas>

T1: Pero, o sea obviamente yo no puedo trabajar todo en el colegio, cachai?

E: Ya.

T1: Yo el fin de semana igual me llevo trabajo para buscar rúbricas diferentes.

E: Ya.

T1: Busco en internet y modifico.

E: Ya.

T1: Eso igual...

E: Claro.

T1: Me ha servido.

E: Eeh, ¿Tú tienes algún conocimiento o tenías un conocimiento previo del diseño de aplicación de rúbricas?

T1: Mm, la verdad es que no. O sea, es como lo que aprendí en la u, súper básico...

E: Ya.

T1: Como no se poh, cinco, tres, uno y ahí yo iba sacando. Ehh... como, me guío harto por lo que sale, por lo que encuentro en en internet.

E: Mmhh.

T1: Algunas sirven, algunas no.

E: Ya.

T1: Yo igual tengo que adecuarlo a mi realidad y a las cosas que yo pido por ejemplo si pido algo específico de vocabulario o de gramática ahí lo voy aplicando....

E: Ya.

T1: Y lo agrego.

E: Perfecto, pero en la u no tuviste así como un ramo de evaluación en donde te enseñaran cómo hacer una rúbrica y todo eso.

T1: Mm no.

E: Ya, super... Ehm...

T1: Solamente metodología, que es como cómo enseñar...

E: Ya, sí.

T1: Pero así como de rúbrica para... no.

E: Y tu me habíai dicho que la rúbrica, eh, te la daban, y que también, ehh... sacabai información...

T1: Sí. me dan algunas, igual te traje unas por si le quieres echar un ojo.

E: Ya, bacán.

T1: Que uno es oral y el otro es escrito.

E: Perfecto.

T1: Sí.

E: Ehm, y esta rúbrica tú la adaptai a las tareas que pedí.

T1: Claro, exacto.

E: Y no solamente, ehm... por ejemplo, por lo general las rúbricas se usan en... en... en cosas orales, ¿Ustedes ocupan solo en presentaciones orales? O también...

T1: No. También en en el oral que se llama “Talking Point”...

E: Mmhh.

T1: Que es una presentación oral de un tema que yo les doy...

E: Mmhh.

T1: Y también en el “Writing Project”...

E: Ya.

T1: Que es para las actividades escritas que se hacen en grupo o individual.

E: Perfect. Ehm, y ¿Cuál es la utilidad que tú le das a la rúbrica? ¿Para que te sirve? ¿Qué, qué demuestra la rúbrica bajo tu percepción?

T1: O sea, primero que todo me sirve para... para poder evaluar y para poner una nota.

E: Ya.

T1: O sea, yo creo que eso es como lo principal y también para que no sea tan… tan... para tratar que la nota sea como lo más justo posible porque por ejemplo, ya, alguien puede hacer un trabajo muy lindo pero si no aplica lo que yo pedí en tema de gramática, de vocabulario, ehh, no sirve.

E: Claro.

T1: Porque puede ser un trabajo bonito pero tampoco esto es arte.

E: Claro.

T1: Entonces por eso me sirve.

E: Mmh, súper. Ehm, y ¿Ocupas varias rúbricas dependiendo de la evaluación que vas a tener?

T1: Claro, por curso... igual son parecidas.

E: Ya.

T1: Voy cambiando… poquitas cosas.

E: Ya, perfecto. Y ¿Quiénes participan en la elaboración de la rúbrica? Dígase las rúbricas que a ti te entregan. ¿Tú sabes las personas que participaron en la creación de esas rúbricas?

T1: Es que mira esas rúbricas ya están hechas hace mucho tiempo.

E: Ya.

T1: Entonces siempre usan las mismas...

E: Ya.

T1: Todos los años.

E: Ya.

T1: Entonces serían los dep... los jefes...

E: Ya

T1: Del departamento de inglés.

E: Y ¿En tu caso para tus cursos tú sola haces las rúbricas o también colaboras con otros profesores en creación de rúbricas o nunca?

T1: No, solo yo porque... eh, a ver... en quinto yo comparto un curso con otra profesora

E: Ya.

T1: Entonces con ella a veces como que compartimos los materiales.

E: Ya.

T1: ¿Ya? Pero en la media como soy la unica lo hago todo yo.

E: Ah ya, perfecto.

T1: Ehm... En, en cuanto a el resultado de tus rúbricas, ¿Se hacen procesos estadísticos para garantizar que todas las notas estén, como, sean válidas... ¿Hacen algún estudio después.. eh media, mediana... o evaluai como tu trabajo respecto a la rúbrica?

T1: Mira, no se hace eso pero yo así personalmente veo y voy revisando los trabajos y, y en las cosas que se... se repiten, por ejemplo si yo veo que en esta sección me faltó o siento que está de más lo modifico para, para la próxima o para los próximos cursos o para el próximo año.

E: Mmhh. ¿Y tu rúbrica se alinea a la tarea que... que tu estas pidiendo? O sea, que los descriptores estén bien hechos...

T1: Sí, sí, sí.

E: Ya.

T1: Generalmente eh, es no se poh “grammar”, “spelling”, “handwriting”, “clean”... eso.

E: Eh, Y, ¿Cuándo llevas a cabo este proceso, o sea en qué momento de, de tu semana eh… tienes esta hora de planificación o te dedicas a, en general a hacer este tipo de rúbricas?

T1: Cuándo lo hago... ehh, bueno... obviamente previo a la... al... a yo como, a explicarle a mis estudiantes lo que vamos a hacer, como una hora antes, o sea disculpa <entre risas>, una semana antes para yo poder hacer la rúbrica o chequearla, mandarla a imprimir y mostrárselo como en ese tiempo.

E: Entonces ehh ¿Tú les muestras a tus alumnos antes la rúbrica?

T1: Claro...

E: Ya.

T1: Sí.

E: Entonces ahí después la ocupai de la misma forma que se la mostraste pa...

T1: Exacto, sí. Porque a veces pasa que yo lo explico y a ellas se le olvida

E: Mm...

T1: ¿Me entiendes? Entonces como que por eso está como “te pasé la rúbrica, échale un ojo”.

E: Ya.

T1: Porque a veces me preguntan “Ya, ¿cuánto vocabulario había que usar?” y yo digo “ahí está”.

E: Mmh.

T1: “Hasta cinco, o desde cinco” entonces ahí es como “ah, verdad”.

E: Ya.

E: Ehm, y antes de las evaluaciones, por ejemplo ya, tenís una presentación oral, tenís una rúbrica, la presentai, ¿la practicas con tus alumnos antes de la evaluación formal? Como una especie de piloteo. Por ejemplo, si vas a evaluar presente simple, les das una guía de presente simple que se va a basar en la prueba por ejemplo.

T1: Ahh ya, ya. Lo que pasa es que, claro, el tema de gramática antes de hacerlo nosotros ya pasamos todas esas unidades. Entonces están súper claras con eso. Si tienen dudas yo les digo pero tampoco es como que les vuelvo a pasar la materia de eso, porque ya lo pasamos.

E: Ya.

T1: Me entiendes? Entonces yo pongo las cosas que voy a evaluar, y ellas ya saben, o la mayoría ya sabe.

E: Ehh... ¿Cuáles son las prácticas respecto al feedback que, que das tú? Por ejemplo subrayas, marcas errores, o corriges altiro....

T1: Mira cuando es oral, ehh... lo voy anotando.

E: Ya.

T1: Por ejemplo si se equivocan dicen, no sé, "we is" pongo “we is” y un círculo y después de, de su evaluar, eh de su presentación, le voy diciendo eh... bueno al comienzo lo hice como una por una, y después lo empecé a hacer en el final de cada clase. De todos los errores más comunes que hay, y entre ellas mismas también yo les pregunto "¿Cuáles crees que son los errores más comunes?" Y ahí me van diciendo y... y lo anotamos en la pizarra.

E: Super, ehh.. y dentro de las prácticas que acabai de mencionar, eh, ¿Cuáles considerarías tú como más efectiva según tu experiencia, como pal’ aprendizaje de tus alumnos?

T1: ¿Con relación al feedback dices tú?

E: Sí, con qué aprenden más.

T1: Ehm, yo siento que aprenden bastante eh al final de cada clase cuando se da el feedback de las cosas que se equivocaron. Entonces la próxima clase o la próxima presentación, están como más ehh... conscientes de que, esas cosas no tiene que hacer, me entiendes?

E: Ya, y este feedback lo das oral y lo haces como en conjunto?

T1: Claro.

E: No lo haces individual.

T1: No, para que todos sepan en lo que se están equivocando, o en lo que están haciendo bien también.

E: Claro. Y ¿Esto lo haces cada clase o casa vez que tienen presentación?

T1: No, cada vez que tienen presentación, que es como un tiempo de... de dos semanas.

E: Ah ya. Ya, y en las evaluaciones escritas, ¿Cuáles son las áreas en las que se enfoca, te enfocas tú al dar tu feedback? Por ejemplo gramática, puntuación, contenido, organización de ideas, etc.

T1: Mira, me enfoco en el contenido que sea de acuerdo, acorde a lo que estamos viendo, gramática, “spelling”...

E: Ya, estamos hablando de las evaluaciones escritas.

T1: <Risas> Eh si, como te decía ehh... me enfoco bastante en, en que el contenido sea de acuerdo a lo que estamos viendo.

E: Mmh.

T1: Ya? Por ejemplo si el tema era eh, de hacer una receta obviamente tiene que ser de eso...

E: Ya.

T1: Me entiendes? A veces no están tan... claros con el contenido.

E: Ya.

T1: Cuando es algo específico.

E: Mmhh.

T1: Entonces ahí yo le bajo puntos.

E: Ya.

T1: Por ejemplo, en primero medio hicimos algo de "¿Qué harías si te pasara esto?" y les di un contexto. Por ejemplo... de conocer gente por internet que no conoces, ¿ya? Listo, pero hay gente que hay, hay chicas, alumnas que, ehm... hablan de otra cosa como de lo mismo pero con animales.

E: Mmhh.

T1: Entonces como, uno queda como… no. <risas> No entendió nada, no estuvo en la clase <entre risas> En el contenido, en la gramática les voy subrayando, en el "spelling" y también en la limpieza del trabajo y calidad de… no, no, no tanto en calidad del trabajo, sino en la limpieza.

E: Ya, si, y ¿Por qué crees que son esos aspectos relevantes a la hora de evaluar?

T1: Mmm.

E: O ¿Qué refleja del alumno por ejemplo? Que el trabajo esté limpio que el trabajo esté ordenado...

T1: Porque demuestra que ehh... si me entrega un trabajo que esté todo doblado, todo sucio, yo voy a entender que ese alumno no se preocupo tanto en... en hacer su trabajo, en mantenerlo limpio, en guardarlo, es más que nada por la… por organización.

E: Mmhh, perfecto.

T1: De saber si esa alumna realmente es ordenada o no.

E: Mmmh, listo, estamos listos.

T1: ¿Eso es todo?

E: Muchas gracias, sí.

Appendix F: Rubrics from Scuola Italiana

Project: "Fabulous food"

The objective of this project is to produce and record a cooking video where you will carry on a research and then explain how to prepare a fabulous dish.

General instructions:

1) The research process will be carried out during classes, from **September 9th to September 12th**.

2) The video must be sent to your teacher's email on **October 4th until midnight**.

Instructions for the fact file:

1) It must be hand written, with no smudges or wrinkles.

2) It must be returned to your teacher **at the end of every class** with **new** information.

Instructions for the video:

1) The video must last between 7 to 10 minutes.

2) The video must include:

- Name of the show
- Introduction of the presenters
- Brief description of the dish
- Ingredients
- Cooking process
- Result and dish tasting

Remember to use:

Cooking tips
Instructions
Catching phrases

"Fabulous food"
Fact File

Names: ______________________________ **Class:** ____________

____

____

Instructions: In order to create this Fact File and your cooking video, you must carry out a research.

- This fact file **must be returned** with your advance at the end of **every class**.
- You must send the list of sources you use every class to the teacher's email at the end of each lesson.

The information that you need to include is the following:

Fact file	
Information of the dish	Name: ______________________________ Country of origin ______________________________
History of the dish	______________________________

Main ingredients of the dish. **- Where do you get them.** **- Can you replace them in case you can't find them?**	
What influences the food on your plate?	

Complete the cooking plan	Dish: ______________________ Ingredients: • ______ ______________________ • ______ ______________________ • ______ ______________________ • ______ ______________________ • ______ ______________________ • ______ ______________________ • ______ ______________________ • ______ ______________________ • ______ ______________________ • ______ ______________________ Cooking instructions: ______________________________ ______________________________ ______________________________ ______________________________ ______________________________ ______________________________ ______________________________ ______________________________ ______________________________ ______________________________ ______________________________ ______________________________ ______________________________ ______________________________ ______________________________ ______________________________ ______________________________

"Fabulous food"

Score: 46 points

Classwork rubric

Classwork	The students behave appropriately every class, respect their classmates and the teacher does not have to remind them that they have to work or to be quiet. They work on the task(s) assigned every day, and present a glossary of new words and interesting ideas gathered from the classes when asked.	**Discount: - 0,1 each time**

Fact file

Classwork (laboratory)	Students work appropriately during the hour in the laboratory; they do not waste their time playing or watching videos that are not related to their research.	2 points
	The students return the fact file every class with new information.	2 points
	The students send all the links of the sources used for their work to the teacher's email at the end of every class.	1 point
Research	The students hand in the fact file with all the required information and supported by reliable sources. (September 10th or 12th at the end of the class)	1 point
	The presentation of the fact file is clean (no smudges or wrinkles) and with a readable handwriting.	1 point
Punctuality	The students send the video to the teacher's email on time: October 4th until midnight.	1 point
Score		8 points

Video rubric

Criteria	Exemplary (3)	Competent (2)	Developing (1)	Score
Narration / Acting / Character Dialogue	• Good pace and innovative use of Narration/Dialogue • Narration/Dialogue uses a wide variety of inflection, pace and emotion • Effective acting.	• Narration/Dialogue has a good pace to match visuals • Emotion and inflection appropriate to on-screen images • Average acting.	• Narrator/Dialogue rushes through or drags behind on screen images • Narration/Dialogue is dry, without emotion or change in inflection. • Poor acting.	__/3
Fact file and planning	• All the important information presented in the fact file and the planning is reflected on the video.	• Some of the important information presented in the fact file and the planning is reflected on the video.	• The important information presented in the fact file and the planning, does not exist on the video	__/3
Story / Message / Information	• The story and message are creative and clearly told. • The story clearly achieves its intent. • Information is organized and presented in a creative manner	• The story and message are somewhat clear. • The story is complete, but it does not achieve its intent. • Information is sequenced in a coherent manner.	• The story and message are not clear. • The story is unorganized, chaotic, or not clearly expressed. • Information is not organized and lacks of coherence.	__/3
Shot Usage	• Shots are well organized. • The purpose and story of the project is clearly discernable.	• Shots indicate a basic organization. • The purpose and story of the project are quite evident.	• Shots are unorganized and do not show planning. • The purpose and story of the project do not exist.	__/3
Sound Clarity	• Little or no unnecessary background noise. • Voices are clear and easy to understand. • Sound is perfectly clear.	• Some unnecessary background noise. • Voices are somewhat clear and easy to understand. • Sound is fairly clear.	• Unplanned background noise. • Hard to hear or understand voices. • Sound is not clear.	__/3
Editing & Special Effects	• Edits are smooth • Edits are creative and enhance the production.	• Edits are somewhat smooth • Edits appropriately link the production.	• Edits lacks continuity • Edits are rough and detract from the flow of the production.	__/3
Score				**__/18**

Speaking rubric

Score	Grammar and vocabulary	Discourse management	Pronunciation	Interactive communication
5	• Shows a good degree of control of a range of simple and some complex grammatical forms. • Uses a range of appropriate vocabulary to give and exchange views on a wide range of familiar topics.	• Produces extended stretches of language with very little hesitation. • Contributions are relevant and there is a clear organization of ideas. • Uses a range of cohesive devices and discourse markers.	• Is intelligible. • Intonation is appropriate. • Sentence and word stress is accurately placed. • Individual sounds are articulated clearly.	• Initiates and responds appropriately, linking contributions to those of other speakers. • Maintains and develops the interaction and negotiates towards an outcome.
4	Performance shares features of bands 3 and 5			
3	• Shows a good degree of control of a range of simple and attempts some complex grammatical forms. • Uses a range of appropriate vocabulary to give and exchange views on a range of familiar topics.	• Produces extended stretches of language despite some hesitation. • Contributions are relevant and there is very little repetition. • Uses a range of cohesive devices .	• Is intelligible. • Intonation is generally appropriate. • Sentence and word stress is accurately placed. • Individual sounds are generally articulated clearly	• Initiates and responds appropriately. • Maintains and develops the interaction and negotiates towards an outcome with very little support.
2	Performance shares features of bands 1 and 3			
1	• Shows a good degree of control of simple grammatical forms. • Uses a range of appropriate vocabulary when talking about everyday situations.	• Produces responses which are extended beyond short phrases, despite hesitation. • Uses basic cohesive devices .	• Is mostly intelligible, and has some control of phonological features at both utterance and word levels.	• Initiates and responds appropriately. • Keeps the interaction going with very little prompting and support.
0	Performance below band 1			
	___/5	___/5	___/5	___/5

Simulazione Colloquio Inglese

Nome :	Punteggio :
Classe:	Voto:

Griglia di valutazione del colloquio

Indicatori			Punteggio candidato
Conoscenze generali e specifiche	Ampie e approfondite	10	
	Complete e precise	9	
	Adeguate	8-7	
	Soddisfacenti	**6**	
	Essenziali	5	
	Frammentarie	4-3	
	Lacunose	2-1	
Capacità elaborative, logiche e critiche	Sintetizza e rielabora i contenuti in modo autonomo e originale	10	
	Sintetizza e rielabora i contenuti in modo personale	9-8	
	Sintetizza e rielabora le proprie conoscenze	7	
	Espone i contenuti e talvolta offre spunti di rielaborazione personale	**6**	
	Espone i contenuti in modo lineare senza operare rielaborazioni personali	5-4-3	
	Espone i contenuti senza operare sintesi e rielabora con difficoltà.	2-1	
Proprietà degli strumenti espressivi	Si esprime sempre in modo chiaro, efficace e corretto sia in italiano che in tutte le lingue straniere	10	
	Si esprime in modo adeguato e preciso.	9-8	
	Si esprime in modo adeguato in italiano commettendo solo occasionalmente alcuni lievi errori.	8-7	
	Si esprime in modo semplice ma sostanzialmente corretto.	**6**	
	Si esprime in modo semplice e vi sono talvolta difficoltà di comunicazione	5-4-3	
	Espressione carente e con difficoltà di comunicazione	2-1	

0	1,0	**5**	1,8	**10**	2,7	**15**	3,5	**20**	4,5	**25**	5,8	**30**	7,0
1	1,2	**6**	2,0	**11**	2,8	**16**	3,7	**21**	4,8	**26**	6,0		
2	1,3	**7**	2,2	**12**	3,0	**17**	3,8	**22**	5,0	**27**	6,3		
3	1,5	**8**	2,3	**13**	3,2	**18**	4,0	**23**	5,3	**28**	6,5		
4	1,7	**9**	2,5	**14**	3,3	**19**	4,3	**24**	5,5	**29**	6,8		

Appendix G: Rubrics from Escuela Arturo Toro Amor

SiP English Department Arturo Toro Amor School Teacher: Da Yon Noh

SiP

TALKING POINT

Name:	Class:	Date:

		Score
Classwork		
1. Uses time appropriately during class. [3 LESSONS]	3 points	
2. Respects classmates' presentations	2 points	
Language		
1. Uses vocabulary appropriately.	2 points	
2. Uses grammar structures appropriately.	2 points	
3. His/her pronunciation is clear and accurate.	4 points	
4. Speaks fluently.	4 points	
Format		
1. Sticks to the time assigned. [1:30 to 3:00 MINUTES]	2 points	
2. Sends the visual material (PPT) to teacher's email on time.	2 points	
3. Use of voice (volume, intonation) and body language (shows confidence, no reading)	2 points	
Content		
1. The information has been organised and presented in the student's own words. (No translator/copy-paste)	3 points	
2. Answers questions appropriately (2)	2 points	
Final Score	28 points	

Comments:______________________________________

Carla Rosas Riquelme
Sub Dirección ATA
SIP Red de Colegios

SiP
RED DE COLEGIOS

ORAL EXAM RUBRIC 1st to 12th GRADE

	Understanding	Use of English	Pronunciation	Communication of Ideas	Bands
100%- 90%	• Understands **without repetition** and/or **prompting.**	• Shows **good control** of simple **grammar** structure. • Uses an **appropriate range of vocabulary** for the level.	• Is **mostly intelligible.** • Has **good** control of **sounds** and **stress.**	• **Answers** questions/keeps **exchanges according** to the **level.** • Explains **ideas clearly.**	5
89%- 85%	Some features of bands 3 and 5				4
84% - 70%	• Understands after **some repetition** and/or **prompting.**	• Shows **sufficient control** of simple **grammar** structures • Show **appropriate vocabulary** for the level. • Makes **a few** elementary **mistakes.**	• Is **mostly intelligible.** • **Has limited** control of **sounds** and **stress.**	• Shows some **difficulty to answer** questions/keep **exchanges according** to the **level.** • Explains **ideas** with **some difficulty.**	3
69% - 45%	Some features of bands 1 and 3				2
44% - 31%	• Understands after **considerable repetition** and/or **prompting.**	• Shows **limited control** of **a few** grammar **structures** and **poor vocabulary** for the level. • Makes **many** elementary **mistakes.**	• **Is mostly intelligible.** • **Has limited** control of **sounds** and **stress.**	• Shows **considerable difficulty** to **answer** questions/keep **exchanges according** to **the level.** • Explains **ideas partially** and with **considerable difficulty.**	1
30% - 10%	Performance below band 1				0

guía

ENGLISH DEPARTMENT

→ relativo

→ sin modificar.

Appendix H: Program for 10th grade in Scuola Italiana

Unidad: 4	**Nombre:** "Fabulous food"		
Asignatura :	Inglés	**Nivel : II**	**Duración: (horas aproximadas) 38**
Descripción de la unidad	En esta unidad los estudiantes podrán explicar cómo hacer un plato, compartir sobre sus experiencias relacionadas a las comidas e ingredientes, discutir situaciones posibles e imaginarias en el presente y futuro, comprender videos y audios donde se describen comidas y escribir un texto no literario donde explican como hacer un plato, sus ingredientes y su historia usando el vocabulario relacionado al tema y estructuras gramaticales tales como first y second conditionals.		
ETAPA I: IDENTIFICAR LOS RESULTADOS DESEADOS			
Ejes : 1) comprensión oral 2) lectura 3) escritura 4) uso del idioma			

CONOCIMIENTOS	**HABILIDADES**	**ACTITUDES**
Textos auditivos: "How to make a dish" Técnicas para discriminar información auditiva. Vocabulario específico de la unidad: Cooking verbs	**OA 1** • Demostrar comprensión de ideas generales e información explícita en textos orales adaptados y auténticos simples, literarios y no literarios, en diversos formatos audiovisuales (como exposiciones orales, conversaciones, descripciones, instrucciones y procedimientos, discursos y debates breves y simples, avisos publicitarios, entrevistas, noticieros, anécdotas, narraciones, canciones), acerca de temas variados (experiencias personales, temas de otras asignaturas, del contexto inmediato, de actualidad e interés global o de otras culturas) y que contienen las funciones del año.	**A.** Manifestar una actitud positiva frente a sí mismo y sus capacidades para aprender y usar el idioma, valorando a la vez, los logros de los demás. **C.** Demostrar interés por el aprendizaje continuo e independiente como parte de un proyecto personal y para contribuir a la sociedad. **E.** Usar de manera responsable y efectiva las tecnologías de la comunicación en la obtención de información y la creación de textos, dando crédito al trabajo de otros y respetando la propiedad y la privacidad de las personas.
Textos auditivos: "Oil from goats?" Técnicas para discriminar información auditiva.	**OA 2** • Identificar palabras, frases y expresiones clave, expresiones idiomáticas, combinaciones frecuentes de palabras (collocations), vocabulario temático, conectores (neither…nor; either…or; though, unless, as soon as, as a result, y los del año anterior), los sonidos iniciales /t/ (to), /d/ (do), sonido / / (usually), combinaciones iniciales /sp/ (special), /st/ (student) y cualidades de la voz (acentuación en palabras de dos o más sílabas y en palabras compuestas) en textos orales en diversos formatos o al participar en interacciones cotidianas y conversaciones en la clase.	
Video temático: "Fruits of the sea?" Textos auditivos: "How to make a dish" Técnicas para identificar ideas generales y específicas dentro de un video o audio. Frases claves para intercambiar información. Vocabulario temático de la unidad.	**OA 3** • Identificar en los textos escuchados: › Propósito o finalidad del texto, tema e ideas relevantes. › Ideas generales en textos sobre temas menos conocidos. › Información específica y detalles clave asociados a personas, sus acciones y opiniones, lugares, tiempo, hablantes y situaciones. › Relaciones de importancia y problema-solución entre ideas.	

	OA 4 • Seleccionar y usar estrategias para apoyar la comprensión de los textos escuchados: › Hacer predicciones. › Escuchar con un propósito. › Usar conocimientos previos. › Hacer inferencias con el apoyo de claves contextuales (tema, participantes, apoyo visual) y pistas fonológicas (entonación, acentuación). › Focalizar la atención en expresiones o frases clave. › Utilizar apoyos como gestos del hablante y entonación. › Pedir repetición o clarificación en interacciones. › Tomar nota de lo escuchado. › Confirmar predicciones.	
Técnicas para presentar e intercambiar información. Expresiones para intercambiar información. Técnicas para plantear temas de discusión.	› Resumir algunas ideas relevantes con apoyo. › Preguntar para confirmar comprensión. **OA 5** • Presentar información en forma oral, usando recursos multimodales que refuercen el mensaje en forma creativa, acerca de temas variados (como experiencias personales, temas de otras asignaturas, otras culturas, problemas globales y textos leídos	
Discusión informal e intercambio de información. Textos auditivos: "A game show"	o escuchados), demostrando: › Conocimiento del contenido y coherencia al organizar ideas. › Uso apropiado de las funciones del lenguaje y vocabulario del nivel. › Uso apropiado de sonidos del idioma, como los sonidos iniciales /t/ (to), /d/ (do), las combinaciones iniciales /sp/ (special), /st/ (student), sonido / / (usually), y cualidades de la voz (acentuación en palabras de dos o más sílabas y en palabras compuestas). › Tener conciencia de audiencia, contexto y propósito. **OA 6** • Participar en interacciones y exposiciones, recurriendo a las siguientes estrategias para expresarse con claridad y fluidez: › Antes de hablar: practicar presentación, organizar la información clave en diagramas, hacer conexiones o establecer relaciones. › Al hablar: parafrasear, usar sinónimos y expresiones de uso frecuente (chunks), usar rellenos temporales (por ejemplo: anyway...; of course; right), activar uso de conectores, autocorregirse con ayuda. › Después de hablar: identificar debilidades, establecer metas, registrar errores y corregirlos con ayuda de recursos.. **OA 7** • Reaccionar a textos leídos o escuchados	

Textos auditivos: "A conversation" Discusión informal e intercambio de información. Vocabulario y gramática de la unidad.	por medio de exposiciones orales o en discusiones y conversaciones grupales en las que: › Hacen conexiones con otras asignaturas, la lengua materna y su cultura, la vida cotidiana, experiencias personales y otras culturas en forma espontánea; por ejemplo: I had read/seen/learned/been; we usually/often… › Evalúan ideas e información; por ejemplo: this is true/correct/false/wrong because… › Expresan sentimientos, interpretaciones o puntos de vista; por ejemplo: I agree/disagree because…; I believe that/in…, if I were you…; I'd rather… › Generan preguntas o hipótesis en forma generalmente correcta; por ejemplo: What would they…? What if…?	
Video temático: "A radio programme" Discusión informal e intercambio de información. Vocabulario relacionado a temas generales e intereses del alumno.	**OA 8** • Demostrar conocimiento y uso del lenguaje en conversaciones, discusiones y exposiciones por medio de las siguientes funciones: › Expresar cantidad, contar y enumerar, por ejemplo: there are a few/few girls; there is little sugar; a large/small number of… › Describir frecuencia, tiempo, grado de acciones, por ejemplo: I usually /often call her in the morning; he is rather shy; Susan will be here soon; he has been working during the summer. › Expresarse con claridad usando palabras y expresiones de uso común, sinónimos, palabras compuestas, por ejemplo: hardly ever; at all; I'd rather…; for ages; I'm pleased; sorry about; look for/after…; cheap enough; online. › Solicitar información sobre frecuencia de actividades, por ejemplo: How often do you practice sports? › Expresar sugerencias, ofrecimientos, hacer solicitudes, por ejemplo: Would you like something to eat?; You ought to talk to him; Shall I bring you a glass of water? › Describir acciones que comenzaron en el pasado y que aún continúan o acaban de finalizar, por ejemplo: we have been training for 3 months; she has been working all day. › Expresar acuerdo o desacuerdo, por ejemplo: I agree/disagree with… › Unir ideas, por ejemplo: neither John nor Susan are in the office; he looks tough, he is shy, though; it is raining today, as a result the air will be cleaner in the city; he told us to go as soon as we could; we wouldn't use gas unless it was cheap. › Expresar énfasis en quien realiza la acción, por ejemplo: Did you make it yourself?; he cut himself. › Describir acciones que ocurrieron antes que otra acción en el pasado, por ejemplo: He didn't have money because he had lost his wallet. › Informar lo que otros dicen, por ejemplo: Mr. Brown said he had finished the report; she told me to go home. › Describir situaciones hipotéticas o imaginarias y dar consejos, por ejemplo: if I were you, I would tell her the truth; he	

Técnicas de comprensión lectora. Video temático: "Giving instructions". Vocabulario y gramática de la unidad.	were you, I would tell her the truth; he would meet more people if he spoke English. **OAC 1** • Identificar mensajes generales y específicos. **OAC 4** • Comprender tipos variados de programas radiales y de televisión, incluyendo boletines informativos y entrevistas, documentales, entrevistas en vivo y programas de conversación. **OAC 6** • Expresar y responder a emociones como sorpresa, felicidad, tristeza, interés o	
Frases comunes para reaccionar a conversaciones en diferentes contextos.	indiferencia. **OAC 7** • Expresar reflexiones en temas más abstractos y culturales como películas, libros, música, etc. **OAC 8** • Dar o buscar opiniones personales y hacer comentarios breves sobre los puntos de vista de otras personas. **OAC 9** • Manifestar de forma respetuosa creencia, opinión, acuerdo y desacuerdo.	
Lectura comprensiva para información y argumentos: - Short online texts Estrategias de comprensión lectora. Vocabulario y gramática de la unidad.	**OAC 12** • Intercambiar, revisar y confirmar información. **OA 9** • Demostrar comprensión de ideas generales e información explícita en textos adaptados y auténticos simples, en formato impreso o digital, acerca de temas variados (como temas de interés de los y las estudiantes, temas de otras asignaturas, del contexto inmediato, de actualidad e interés global, de otras culturas y algunos temas menos conocidos) y que contienen las funciones del año. **OA 10** • Demostrar comprensión de textos no literarios (como descripciones, artículos de revistas, procedimientos, catálogos, avisos publicitarios, emails, diálogos, páginas web, biografías, gráficos, reseñas, noticias, discursos) al identificar: › Propósito del texto. › Idea principal, información específica y detalles. › Relaciones de importancia y problema-	
	solución entre ideas. › Palabras y frases clave, expresiones idiomáticas y de uso frecuente y vocabulario temático.	

	› Combinaciones frecuentes de palabras (collocations), conectores, (neither… nor; either…or; though, unless, as soon as, as a result, y los del año anterior) y palabras derivadas de otras por medio de los prefijos im-, in-, il-.	
Estrategias para leer comprensivamente y discriminar información relevante de un texto.	**OA 12** • Seleccionar y usar estrategias para apoyar la comprensión de los textos leídos: › Prelectura: leer con un propósito, usar conocimientos previos. › Lectura: hacer lectura rápida y lectura focalizada, hacer inferencias, releer, identificar elementos organizacionales del texto (título, subtítulo, diagramas), plantearse preguntas al leer. › Postlectura: responder preguntas planteadas, usar organizadores gráficos, releer, recontar, resumir, preguntar para confirmar información.	
Estrategias de comprensión lectora, para extraer la idea general e información específica de un texto escrito.	**OAC 20** • Identificar las conclusiones principales dentro de textos argumentativos claramente señalados. **OAC 21** • Reconocer la línea de argumento dentro de un tema, aunque no necesariamente en detalle. **OAC 22** • Reconocer aspectos importantes y obtener información, ideas y opiniones de un artículo claro ligado a temas cercanos.	
Técnicas de escritura. Vocabulario y gramática de la unidad. Estructura de texto: artículo	**OA 14** • Escribir una variedad de textos, como cuentos, correos electrónicos, artículos, folletos, rimas, descripciones, biografías, instrucciones, cartas, resúmenes, diario personal (journal) y poemas o canciones utilizando los pasos del proceso de escritura (organizar ideas, redactar, revisar, editar, publicar), generalmente en forma independiente de acuerdo a un criterio de evaluación, recurriendo a herramientas como el procesador de textos y diccionario en línea.	
Introducción de ideas y argumentos con coherencia. Producción general escrita. Vocabulario y gramática de la unidad.	**OA 15** • Escribir para analizar, expresar opiniones y narrar, usando: › Variedad de palabras, oraciones y estructuras aprendidas y de otros contextos. › Conectores variados. › Correcta ortografía de palabras de uso frecuente en forma consistente. › Puntuación apropiada (comillas).	
Modelos de texto. -Vocabulario de la unidad: -Gramática de la unidad.	**OA 16** • Demostrar conocimiento y uso del lenguaje en sus textos escritos por medio de las siguientes funciones: › Expresar cantidades, contar y enumerar; por ejemplo: there are a few/few girls; there is little sugar left. › Describir frecuencia, tiempo, grado de acciones; por ejemplo: I usually /often call her in the morning; he is rather shy; Susan will be here soon. › Describir personas y situaciones; por ejemplo: he is polite/impolite; it was complete/incomplete/legal/illegal. › Expresarse con claridad usando palabras y expresiones de uso común, sinónimos y palabras compuestas; por ejemplo: enough	

	palabras compuestas; por ejemplo: enough time; sorry about, belong to, look for…; online. › Solicitar información sobre frecuencia de actividades; por ejemplo: How often do you practice sports?	
	› Describir acciones que comenzaron en el pasado y que aún continúan o acaban de finalizar; por ejemplo: we have been training for 3 months; she has been working all day. › Expresar acuerdo o desacuerdo, opinion, posesión, tiempo; por ejemplo: I agree/disagree with…; I was left without…; he has been working during the summer; I believe that/in… › Unir ideas; por ejemplo: neither John nor Susan are in the office; he looks tough, he is shy, though; It is raining today, as a result the air will be cleaner in the city; he told us to go as soon as we could. › Expresar énfasis en quien realiza la acción; por ejemplo: he cut himself. › Describir acciones que ocurrieron antes de otra acción en el pasado; por ejemplo: He didn't have money because he had lost his wallet. › Informar lo que otros dicen; por ejemplo: Mr. Brown said he had finished the report; she told me to go home. › Describir situaciones hipotéticas o imaginarias y dar consejos; por ejemplo: if I were you, I would tell her the truth; he would meet more people if he spoke English.	
Vocabulario relacionado a temas generales e intereses del alumno. Modelo de texto argumentativo.	**OAC 23** • Escribir textos claros conectados con un rango de temas habituales dentro de sus intereses, uniendo una serie de elementos dentro de una secuencia lineal. **OAC 25** • Escribir una descripción de un evento, un viaje reciente (real o imaginario). **OAC 27** • Narrar una historia. **OAC 29** • Usar una variedad de "linking words" y	
Connectors and sequencers para introducir argumentos con coherencia Uso: Parts of speech. Presentación oral. Técnicas y expresiones para presentar e intercambiar información.	elementos cohesivos de una manera eficiente. **OAC 36** • Poseer conciencia de las principales normas de cortesía y actuar adecuadamente. **OAC 41** • Formular preguntas complementarias para corroborar la correcta comprensión del mensaje, y aclarar los puntos ambiguos.	

Técnicas y expresiones para presentar e intercambiar información.	mensaje, y aclarar los puntos ambiguos.	

ETAPA II: EVIDENCIA DEL APRENDIZAJE DE LOS ALUMNOS	
Criterios de desempeño: Se expresa a través de textos escritos y discursos orales gramaticalmente correctos, donde de instrucciones claras y precisas de cómo hacer algo y discutir situaciones imaginarias, demostrando respeto, interés y tolerancia por otras culturas, la propia y valorando su aporte al conocimiento.	**Indicadores de logro:** - Lee comprensivamente textos acerca de logros personales. - Escucha y discrimina información de textos orales y escritos relacionados a planes futuros. - Identifica y utiliza el vocabulario y los conocimientos morfosintácticos de la unidad para discutir opciones y tomar decisiones. - Utiliza sus experiencias personales e investiga para redactar un ensayo de opinión. - Prepara y expone sobre un personaje destacado.

Ejemplos de evaluaciones:

Sumativa: - Presentación oral "Fabulous food".

Tareas de desempeño: Los alumnos utilizan los conocimientos lexicales y gramaticales para presentar un tema relacionado a la unidad.

Formativa:

- Trabajo en laboratorio retroalimentado por el profesor. Preguntas abiertas, formuladas para invitarles a explorar sus ideas y sus razonamientos.
- Interrogaciones orales para chequear comprensión y adquisición de léxico y estructuras gramaticales.
- Producción de artículo guiado y retroalimentado por el profesor.

ETAPA III: PLANIFICACIÓN DE LAS EXPERIENCIAS DE APRENDIZAJE

Ejemplos experiencias de Aprendizaje a desarrollar	Recursos pedagógicos

Ejemplos de evaluaciones:

Sumativa: - Presentación oral "Fabulous food".

Tareas de desempeño: Los alumnos utilizan los conocimientos lexicales y gramaticales para presentar un tema relacionado a la unidad.

Formativa:

- Trabajo en laboratorio retroalimentado por el profesor. Preguntas abiertas, formuladas para invitarles a explorar sus ideas y sus razonamientos.
- Interrogaciones orales para chequear comprensión y adquisición de léxico y estructuras gramaticales.
- Producción de artículo guiado y retroalimentado por el profesor.

ETAPA III: PLANIFICACIÓN DE LAS EXPERIENCIAS DE APRENDIZAJE

Ejemplos experiencias de Aprendizaje a desarrollar	Recursos pedagógicos

Ejemplo de actividades	
- análisis de textos, audios y videos. - identificación y practica de vocabulario especifico y estructuras gramaticales en workbook. - discusiones grupales acerca de planes para el futuro y donar dinero. - interrogaciones orales. - práctica de ensayo de opinión con modelo. - trabajo en laboratorio de computación para practicar y reforzar los contenidos vistos en la unidad.	- Texto: "Eyes open 4A"(Student's book and workbook) - Diccionario - Computador - Data - laboratorio de computación. -www.cambridgelms.org/main (Online practice)

Appendix I: Program for 10th grade in Escuela Arturo Toro Amor

10TH GRADE	**EYES OPEN 4A**	Review Unit Test 1: Unit “Trends” part 1 Used to, would, clothes, adjectives. Test 2: Unit “Trends” part 2 Past perfect, dependent prepositions, compound nouns. Final Test: Unit “A helping hand” Reflexive pronouns, each other, present perfect simple vs. present continuous, personal qualities, word building, phrasal verbs	Test 1: Unit “Young achievers” part 1 Be going to and present perfect for the future, training and qualifications, Test 2: Unit “Young achievers” part 2 Predictions with be going to, will, may, might, future continuous, expressions with take, achievements. Final Test: Unit “Fabulous food” 1st conditional when/unless, second conditional could/might, cooking verbs, adjectives describing food, prepositional phrases.

Appendix J: Meta-Rubric

DIMENSIÓN	SUBDIMENSIONES	CONSTRUCTO	FUENTES DE VERIFICACIÓN
CONSTRUCTO	Teórico	Consistente con el conocimiento actuar del desempeño de lo que se está evaluando	Indicador: Dimensions/ Descriptor
	Pedagógico	Que se ajuste con los objetivos y contenidos del curso	Indicador: Dimensions/ Descriptor (Habría que buscar cuales son los enfoques de esta dimensión, deberían aparecer en el programa del curso, y si no están, no deberían estar en la rúbrica)
	Task-related	Que responda adecuadamente a la tarea que se está pidiendo / limitado a los aspectos del constructo	Indicador: Descriptor/ Specimen-Register

DIMENSIÓN	SUBDIMENSIONES	CONSTRUCTO	FUENTES DE VERIFICACIÓN
PROCEDIMIENTOS	Calibración intra/inter evaluador	Sirve para la confiabilidad (es externo a la rúbrica en tanto al contexto, los evaluadores se entrenan para calibrar la confiabilidad entre los evaluadores- que todos usen la misma rubrica y que todos pongan la misma rúbrica)	
	Evaluación continua	Donde la rúbrica está describiendo bien y donde está describiendo mal	
	Proceso de diseño	Mecanismos de adaptación/ creación	
		Base empírica: Ojalá que la rúbrica se base en los desempeños observable que se quiere evaluar	
	Calificación/ Asignación de puntaje	Decidir cuándo evaluar	
		Mecanismos de acuerdo: arbitraje, conciliación, promedio, ponderaciones, etc.	
		Mecanismos de cálculo de nota/puntaje: traducción de rúbrica a escala.	
	Feedback	Afecta la confiabilidad de la rúbrica y la validez del proceso de evaluación	