

Beurteilung der Sprechkompetenz im DaF-Unterricht:

Praktiken und Ansichten von Lehrpersonen und Schüler/innen der Sekundarstufe I

Évaluation de la production orale en classe d'allemand langue étrangère:

pratiques et points de vue des enseignant·e·s et des élèves du secondaire I

Valutazione della produzione orale nelle classi di tedesco come lingua straniera:

pratiche e opinioni di insegnanti e allieve/i di scuola media

Assessment of speaking skills in the German as a foreign language classroom:

practices and views of lower secondary school teachers and students

Elisabeth Peyer, Gabriela Lüthi, Nadia Ravazzini

2025

Bericht des Wissenschaftlichen Kompetenzzentrums für Mehrsprachigkeit

Rapport du Centre scientifique de compétence sur le plurilinguisme

Rapporto del Centro scientifico di competenza per il plurilinguismo

Report of the Research Centre on Multilingualism

Publié par | Herausgeber
Institut de plurilinguisme
www.institut-plurilinguisme.ch

—
Institut für Mehrsprachigkeit
www.institut-mehrsprachigkeit.ch

Auteur-e-s | Autor*innen
Elisabeth Peyer, Gabriela Lüthi, Nadia Ravazzini
Supervision: Thomas Studer

Le projet dont il est question a été réalisé dans le cadre du programme de recherche 2021-2024 du Centre scientifique de compétence sur le plurilinguisme. La responsabilité du contenu de la présente publication incombe à ses auteur-e-s.


Das vorliegende Projekt wurde im Rahmen des Forschungsprogramms 2021-2024 des Wissenschaftlichen Kompetenzzentrums für Mehrsprachigkeit durchgeführt. Für den Inhalt dieser Veröffentlichung sind die Autor*innen verantwortlich.

Übersetzungen | Traductions
Joël Rey – Traduzioni e redazioni, pro-verbial, tran-scribe (Mary Carozza)

Freiburg | Fribourg, 2025

Layout
Billy Ben, Graphic Design Studio

Unterstützt von | avec le soutien de

 Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI
Département fédéral de l'intérieur DFI
Dipartimento federale dell'interno DFI
Departament federal da l'interno DFI
Bundesamt für Kultur BAK
Office fédéral de la culture OFC
Ufficio federale della cultura UFC
Uffizi federal da cultura UFC

Beurteilung der Sprechkompetenz im DaF-Unterricht:

Praktiken und Ansichten von Lehrpersonen und Schüler/innen der Sekundarstufe I

Évaluation de la production orale en classe d'allemand langue étrangère:

pratiques et points de vue des enseignant-e-s et des élèves du secondaire I

Valutazione della produzione orale nelle classi di tedesco come lingua straniera:

pratiche e opinioni di insegnanti e allieve/i di scuola media

Assessment of speaking skills in the German as a foreign language classroom:

practices and views of lower secondary school teachers and students

—
Elisabeth Peyer, Gabriela Lüthi, Nadia Ravazzini

2025
Bericht des Wissenschaftlichen Kompetenzzentrums für Mehrsprachigkeit
Rapport du Centre scientifique de compétence sur le plurilinguisme
Rapporto del Centro scientifico di competenza per il plurilinguismo
Report of the Research Centre on Multilingualism

Inhalt

Deutsch	7
1 Ausgangslage	8
2 Forschungsfragen	10
3 Projektphasen und Forschungsmethoden	11
4 Resultate	14
4.1 Welche formativen Beurteilungspraktiken wenden Lehrpersonen auf Sekundarstufe I an und wie begründen sie diese?	14
4.2 Wie werden Paarprüfungen auf der Sek I durchgeführt?	15
4.3 Welchen Nutzen sehen Lehrpersonen in der summativen Beurteilung des Sprechens?	21
4.4 Welche Bedürfnisse haben die Lehrpersonen bzgl. der summativen Beurteilung des Sprechens?	23
4.5 Was denken die Schüler/innen über Sprechprüfungen?	24
5 Fazit	27
6 Bibliografie	107
7 Transkriptionszeichen	111

Français	31
1 Situation initiale	32
2 Questions de recherche	34
3 Phases du projet et méthodologie de recherche	35
4 Résultats	38
4.1 Quelles sont les pratiques d'évaluation formative utilisées par les enseignant-e-s du secondaire I et comment sont-elles explicitées?	38
4.2 Comment les épreuves orales par deux se déroulent-elles au niveau secondaire I?	39
4.3 Quelle est l'utilité que les enseignant-e-s voient dans l'évaluation sommative de l'expression orale?	46
4.4 Quels sont les besoins des enseignant-e-s en matière d'évaluation sommative de l'expression orale?	48
4.5 Que pensent les élèves des épreuves d'expression orale?	49
5 Conclusion	52
6 Bibliographie	107
7 Conventions de transcription	111

Italiano	57
1 Situazione di partenza	58
2 Domande di ricerca	59
3 Fasi del progetto e metodi di ricerca	60
4 Risultati	63
4.1 Quali pratiche di valutazione formative adottano le e gli insegnanti del livello secondario I e come motivano questa scelta?	63
4.2 Come vengono svolte le prove a coppie al livello secondario I?	64
4.3 Quale utilità vedono le e gli insegnanti nella valutazione sommativa della produzione orale?	71
4.4 Quali esigenze hanno le e gli insegnanti per la valutazione sommativa del parlato?	72
4.5 Che cosa pensano le allieve e gli allievi delle prove orali?	74
5 Conclusione	77
6 Bibliografia	107
7 Convenzioni di trascrizione	111

English	81
1 Starting point	82
2 Research questions	84
3 Project phases and research methods	85
4 Findings	88
4.1 What formative assessment practices do lower secondary school teachers use and how do they explain their decisions?	88
4.2 How are paired speaking tests conducted at the lower secondary level?	89
4.3 What benefits do teachers identify in summative assessments of speaking skills?	96
4.4 What needs do teachers have regarding the summative assessment of speaking skills?	97
4.5 What do students think about speaking tests?	99
5 Final remarks	102
6 Bibliography	107
7 Transcription conventions	111

Beurteilung der Sprechkompetenz im DaF-Unterricht:

Praktiken und Ansichten von Lehrpersonen und Schüler/innen
der Sekundarstufe I

—

Elisabeth Peyer, Gabriela Lüthi, Nadia Ravazzini

1 Ausgangslage

Beurteilungen mündlicher Leistungen – von der Reaktion auf Schülermeldungen über die Rückmeldung zu Präsentationen bis hin zu Prüfungen – gehören zum Unterrichtsalltag und spielen eine zentrale Rolle im Lehr-Lernprozess. So zeigen Beurteilungsaktivitäten den Schüler/innen und Lehrpersonen das bisher Gelernte bzw. noch bestehende Lücken auf. Zudem bieten sie den Lehrpersonen idealerweise Informationen darüber, wie sie noch besser auf die Schüler/innen eingehen können, damit diese die Lernziele erreichen. Verschiedene Befragungen von Fremdsprachlehrpersonen haben jedoch ergeben, dass sich Lehrpersonen bei Beurteilungsfragen tendenziell unsicher fühlen und sich mehr Kenntnisse und praktisches Wissen zu verschiedenen Fragen der Beurteilung, besonders auch zur Beurteilung des Sprechens wünschen (u. a. Berry et al., 2019; Vogt, 2018). Vor diesem Hintergrund interessiert sich die vorliegende Studie für die so genannte *Language Assessment Literacy* von Lehrpersonen bezüglich der Sprechkompetenz. Im untersuchten Kontext – Unterricht von Deutsch als Fremdsprache auf Sekundarstufe I in der französisch- und italienischsprachigen Schweiz – ist die *Language Assessment Literacy* der Lehrpersonen von besonderer Bedeutung, da Deutsch obligatorisch ist und die Deutschnote für die weitere Schullaufbahn der Jugendlichen ausschlaggebend sein kann. Unter der *Language Assessment Literacy* versteht man nebst den notwendigen Kenntnissen und Fähigkeiten, Beurteilungsanlässe zu konzipieren und durchzuführen sowie Leistungen zu bewerten auch das Bewusstsein für die Auswirkungen, die Beurteilungen auf den Einzelnen und die Gesellschaft insgesamt haben können (u. a. Inbar-Lourie, 2008; Fulcher, 2012; Vogt & Tsagari, 2014). So betont z. B. Inbar-Lourie, dass zur (*Language*) *Assessment Literacy* auch die Fähigkeit gehört, kritische Fragen zum Zweck der Beurteilung, zur Eignung des verwendeten Instruments und zu den Testbedingungen zu stellen sowie zu wissen, was auf der Grundlage der Testergebnisse geschehen wird (2008: 389).

Mit der Überprüfung der mündlichen Kompetenz gehen verschiedene Schwierigkeiten einher. So ist die Beurteilung der Mündlichkeit aufgrund ihrer Flüchtigkeit schwierig und es können im Beurteilungsprozess verschiedene Faktoren vorkommen, die keinen Einfluss auf die Leistungsbeurteilung haben sollten (konstruktirrelevante Varianz), wie Aufgaben, Kriterienraster, Beurteiler/innen oder auch Gesprächspartner/innen (Isaacs, 2016; De Jong, 2023). Für eine valide Beurteilung der Sprechkompetenz brauchen Lehrpersonen gemäss Csépes & Fekete deshalb insbesondere ein gutes Verständnis der Grundprinzipien der Aufgabenentwicklung sowie vom Beurteilungsvorgang und von Kriterienrastern (2018: 99). Zudem sind für eine kriterienorientierte Beurteilung mithilfe eines Rasters eine Trainingsphase sowie illustrative Beispiele nötig (ebd.). Das summative Assessment der Sprechkompetenz in der Schule ist auch deshalb eine Herausforderung, da es einerseits zeitaufwändig ist und andererseits für die Lehrpersonen eine besonders komplexe Aufgabe darstellt: So müssen sie gleichzeitig mehrere Rollen übernehmen, die z. B. bei standardisierten internationalen Prüfungen normalerweise auf mehrere Personen verteilt sind, wie

Testleiter/in, Beurteiler/in und je nach Testformat auch Gesprächspartner/in (Teasdale & Leung, 2000; Koizumi, 2022). Gleichzeitig sind sie immer auch noch Lehrperson der betreffenden Schüler/innen.

2 Forschungsfragen

Vor dem oben beschriebenen Hintergrund untersuchte das Projekt *Beurteilung der Sprechkompetenz im DaF-Unterricht*, welche formativen und summativen Beurteilungspraktiken Lehrpersonen von Deutsch als Fremdsprache (DaF) auf der Sekundarstufe I in der Schweiz in Bezug auf die Sprechkompetenz anwenden und welche Ansichten sie diesbezüglich haben. Während bereits verschiedentlich Befragungen von Lehrpersonen zu ihren Beurteilungspraktiken, ihrer Ausbildung und allfälligen Weiterbildungsbedürfnissen im Bereich Language Testing und Assessment durchgeführt wurden (u. a. Berry et al., 2019; Fulcher, 2012; Tsagari & Vogt, 2017; Vogt & Tsagari, 2014; Vogt, 2018), gibt es – gerade zum obligatorischen Schulkontext – noch kaum Studien, die Beurteilungspraktiken beobachten und diese mit den erfragten Ansichten der Lehrpersonen in Verbindung bringen. Dies ist umso erstaunlicher, da Beurteilungspraktiken nicht direkt erfragt werden können und mit den Erfahrungen und Einstellungen der Lehrpersonen zusammenhängen (Berry et al., 2019; Scardino, 2013). Da die Ansichten und Wünsche der direktbetroffenen Lernenden zu Fragen des Assessments bis jetzt in der Forschung ebenfalls kaum Beachtung gefunden haben, sollten auch diese in unserer Studie erfasst werden. Letztlich ging es uns darum, herauszufinden, wie Lehrpersonen und auch Schüler/innen in Bezug auf die Beurteilung der Sprechkompetenz bestmöglich unterstützt werden können.

Folgende Forschungsfragen wurden gestellt:

1. Welche formativen Beurteilungspraktiken wenden DaF-Lehrpersonen in Bezug auf das Sprechen an und wie begründen sie ihre Praktiken?
2. Welchen Nutzen sehen die Lehrpersonen in der summativen Beurteilung des Sprechens und welche Bedürfnisse (nach Hilfestellungen, Materialien) haben sie diesbezüglich?
3. Wie werden summative Beurteilungen der Sprechkompetenz durchgeführt?
 - a. Wann und wie intervenieren Lehrpersonen während Paarprüfungen?
 - b. Welche Auswirkungen haben die Interventionen der Lehrpersonen auf den weiteren Verlauf der Peer-Interaktion?
4. Welche Ansichten und Wünsche äussern die Schüler/innen bzgl. der Beurteilung des Sprechens?

Frage 1 sollte aufgrund von Unterrichtsbeobachtungen und retrospektiven Lehrpersoneninterviews beantwortet werden, Frage 2 ebenfalls aufgrund von Lehrpersoneninterviews. Fragekomplex 3 wurde mithilfe von videographierten authentischen Sprechprüfungen untersucht und Frage 4 mithilfe eines Schüler/innen-Fragebogens und retrospektiven Interviews mit den gefilmten Schüler/innen.

3 Projektphasen und Forschungsmethoden

In einer ersten Projektphase lag der Schwerpunkt auf dem formativen Assessment. Um zu untersuchen, welche formativen Beurteilungspraktiken die Lehrpersonen einsetzen und welche Einstellung zur Beurteilung die Lehrpersonen haben, wurden in zwölf Klassen der Sekundarstufe I jeweils zwei DaF-Lektionen beobachtet und anschliessend mit den Lehrpersonen Interviews geführt.¹ An der Studie teilgenommen haben acht Lehrpersonen mit Klassen des unteren und mittleren Anforderungsniveaus der 8. und 9. Klassenstufe des Kantons Freiburg (10H und 11H nach HarmoS) sowie vier Klassen aus dem Kanton Neuenburg.² Die Lehrpersonen verfügten alle über eine Ausbildung für den DaF-Unterricht auf Sekundarstufe I. Ihre Unterrichtserfahrung in DaF auf Sek-I-Niveau lag zwischen zwei und 25 Jahren und betrug im Mittel 7.75 Jahre. Alle Lehrpersonen haben freiwillig an der Studie teilgenommen. Es dürfte sich also um Lehrpersonen handeln, die ein Interesse am Thema „(formative) Beurteilung des Sprechens“ haben und sich in diesem Bereich einigermaßen sicher fühlen. Da ein möglichst authentischer Unterricht beobachtet werden sollte, wurden die Lehrpersonen lediglich gebeten, Lektionen „mit Fokus auf dem Sprechen“ durchzuführen. Das Lehrpersonen-Interview war teilstandardisiert und wurde im Anschluss an die zweite Lektion durchgeführt. Es diente dazu, die Ansichten der Lehrpersonen zur Beurteilung der Sprechfertigkeit zu erfassen, dauerte jeweils ca. eine Stunde und bestand aus zwei Teilen. Im ersten Teil wurde die Lehrperson auf ca. drei Unterrichtssituationen angesprochen, bei denen Beurteilungspraktiken beobachtet worden waren. Nachdem die Interviewerin eine Situation skizziert hatte, erklärten die Lehrpersonen zumeist direkt, warum sie dies gemacht hätten und in welchen Situationen sie diese Praktik anwenden würden. Im zweiten Teil des Interviews wurden weitere Praktiken und Ansichten der Lehrpersonen zur (formativen und summativen) Beurteilung des Sprechens erfasst. Die Interviews wurden mithilfe des GAT-2-Minimaltranskripts transkribiert (Selting et al., 2009) und danach inhaltsanalytisch ausgewertet (Mayring, 2010).

In der zweiten Projektphase wurde der Fokus auf summative Sprechprüfungen gelegt. In acht Sekundarklassen³ wurden authentische Sprechprüfungen gefilmt und anschliessend mit den Lehrpersonen und Schüler/innen retrospektive Interviews durchgeführt. Darüber hinaus füllten die Schüler/innen einen Fragebogen zu Sprechprüfungen im Fach Deutsch als Fremdsprache aus (für Details zu den Erhebungsinstrumenten vgl. Abb. 1). Alle Prüfungen wurden

- 1 Wir möchten uns an dieser Stelle bei allen Lehrpersonen und Schüler/innen, die an der Studie teilgenommen haben für ihr Vertrauen und ihre Offenheit herzlich bedanken. Ein grosses Dankeschön geht auch an Thomas Studer sowie unsere wissenschaftliche Begleitgruppe für ihre zahlreichen konstruktiven Rückmeldungen zu den verschiedenen Projektetappen.
- 2 Im Kanton Neuenburg werden in der 7. Klasse (9H nach der HarmoS-Zählweise) im Deutschunterricht noch keine Anforderungsniveaus unterschieden. In der 8. und 9. Klasse (10H und 11H) wird Deutsch dann in zwei Niveaus unterrichtet. Beobachtet wurden in Neuenburg eine 7. Klasse sowie zwei Klassen mit Grund- und eine mit erweiterten Anforderungen.

von den Lehrpersonen durchgeführt und fanden auf unseren Wunsch zumeist in einem separaten Raum statt. Die Kamera wurde so aufgestellt, dass Schüler/innen und Lehrperson auf dem Film zu sehen sind. Während der Prüfungen war das Forschungsteam nicht im Prüfungsraum anwesend, sondern führte die schriftliche Schülerbefragung und die retrospektiven Interviews mit den Schüler/innen durch. Die retrospektiven Interviews mit den Lehrpersonen wurden später am selben Tag oder am nächsten Tag online über MS Teams durchgeführt.⁴

Im Anschluss an die Datenerhebungen wurden die Sprechprüfungen und Interviews mittels des GAT-2-Minimaltranskripts transkribiert, wobei die Transkripte der Prüfungen auch Gestik und Mimik erfassen. Die darauffolgende Konversationsanalyse fokussierte auf das Interventionsverhalten der Lehrpersonen bei Paarprüfungen, da bei diesem Punkt eine grosse Variabilität augenfällig war (für Details vgl. Peyer et al., eingereicht). Folgende Forschungsfragen kristallisierten sich als wesentlich heraus:

1. Wann und wie intervenieren Lehrpersonen auf Sekundarstufe I bei mündlichen Paarprüfungen?
2. Welchen Effekt haben verschiedene Typen von Interventionen auf die Peer-Interaktion?

Zur Beantwortung der zweiten Frage stützen wir uns auf das Konzept der *Progressivity* (Schegloff, 2007), womit das Fortschreiten des Gesprächs gemeint ist. Dabei operationalisieren wir *Progressivity* wie folgt:

- a. Interventionen von Lehrpersonen sind förderlich für das Fortschreiten der Peer-Interaktion, wenn diese nach der Intervention ohne eine weitere verbale Intervention der Lehrperson fortgesetzt werden kann.
- b. Interventionen sind für das Fortschreiten der Peer-Interaktion wenig förderlich, je mehr weitere verbale Interventionen der Lehrperson notwendig sind, bevor die Peer-Interaktion fortgesetzt werden kann oder wenn die Kandidat/innen Nachfragen stellen müssen.

Die Interviews mit den Lehrpersonen und den Schüler/innen wurden inhaltsanalytisch ausgewertet (Mayring, 2010). Der Schüler/innen-Fragebogen bestand aus geschlossenen und offenen Fragen. Die offenen Fragen wurden ebenfalls inhaltsanalytisch ausgewertet (für Details zur Befragung der Schüler/innen vgl. Lüthi et al., eingereicht).

3 Im Kanton Freiburg haben vier Klassen teilgenommen: drei Klassen des untersten Anforderungsniveaus (zwei 8. Klassen (10H) und eine 9. Klasse (11H)) sowie eine 8. Klasse (10H) des mittleren Anforderungsniveaus. Aus dem Kanton Neuenburg haben zwei Klassen teilgenommen: eine 7. Klasse (9H, ohne Niveau-Zuteilung) und eine 9. Klasse (11H) des tieferen Niveaus. Aus dem Kanton Tessin haben zwei Klassen teilgenommen: eine 8. Klasse (10H) des höheren und eine 9. Klasse (11H) des tieferen Anforderungsniveaus.

4 Die Erhebungsinstrumente der Studie sind zugänglich unter: <https://osf.io/vwuf2/>

	Methode	Teilnehmer/innen
Projektphase 1 März-Juni 2021	<i>Unterrichtsbeobachtung</i> Protokollieren des formativen Assessments, insbesondere Rückmeldungen der Lehrpersonen	12 Klassen/Lehrpersonen à 2 Lektionen (8 aus Freiburg, 4 aus Neuenburg)
	<i>Retrospektive Interviews mit Lehrpersonen</i> Teil 1 (retrospektiv) zu zuvor beobachteten Beurteilungs-Situationen, Teil 2: allgemeine Fragen zur Beurteilung	12 Lehrpersonen
Projektphase 2 März-Juni 2022	<i>Videografie authentischer Sprechprüfungen</i> von der Lehrperson erstellte und durchgeführte Prüfungen (Paarprüfungen und Einzelprüfungen)	82 SuS: 48 Prüfungen in 8 Klassen (4 aus Freiburg, 2 aus Neuenburg, 2 aus dem Tessin)
	<i>Retrospektive Interviews mit Lehrpersonen</i> (Themen: Einschätzung der Prüfung/Leistung der Klasse, Vorbereitung, Beurteilung/Notengebung, Nützlichkeit mündlicher Prüfungen)	7 Lehrpersonen (4 aus Freiburg, 2 aus Neuenburg, 1 aus dem Tessin mit 2 Klassen)
	<i>Retrospektive Interviews mit Schüler/innen</i> Kurze Interviews mit gefilmten Schüler/innen (Themen: Selbsteinschätzung, Vorbereitung, Nützlichkeit)	Interviews mit 84 SuS
	<i>Schüler/innen-Fragebogen</i> (Themen: Prüfungsvorbereitung, Prüfungsangst, Feedback, Nützlichkeit verschiedener Prüfungsformate, Selbsteinschätzung)	254 SuS (17 Klassen) (Freiburg: 57%, Neuenburg: 13%, Tessin: 30%)

Abbildung 1: Überblick über die Datenerhebungen

4 Resultate

Im Folgenden werden die wichtigsten Resultate der zwei Projektphasen zusammenfassend dargestellt.

4.1 Welche formativen Beurteilungspraktiken wenden Lehrpersonen auf Sekundarstufe I an und wie begründen sie diese?

In den 24 beobachteten Lektionen führten die Lehrpersonen mehrheitlich strukturierte Sprechaktivitäten durch (Arbeit mit Dialoggeländern, Fragen beantworten, Spiele), die als primäres Ziel das Üben neuer Wörter und grammatischer Strukturen verfolgten. Es zeigte sich, dass die Lehrpersonen dabei v.a. spontanes korrektives Feedback geben. Geplante, formellere formative Beurteilungs- bzw. Feedbackpraktiken konnten nur selten beobachtet werden. Die häufigste beobachtete Form des korrektiven Feedbacks waren *Recasts*, d.h. korrekte Wiederholungen von fehlerhaften Äusserungen durch die Lehrperson (Lyster & Ranta, 1997). *Recasts* werden, gemäss Aussagen in den Interviews, von manchen Lehrpersonen gerade darum geschätzt, weil sie nicht explizit sagen müssen, dass ein Fehler vorliegt, sondern – so die Annahme von zwei Lehrpersonen – die Lernenden den Fehler selber bemerken würden. Auch *Prompts* (verschiedene Hinweise, die zur Selbstkorrektur anregen Lyster & Saito, 2010) wurden von den Lehrpersonen regelmässig, aber nicht häufig eingesetzt. Bei einigen wenigen Klassen konnten während einer Sprechaktivität auch Peer-Korrekturen und -Erklärungen beobachtet werden. Eine Lehrperson erklärte im Interview, die Peer-Korrektur bewusst zu fördern, weil die Schülerinnen und Schüler eine andere Art zu erklären hätten, und sie den Eindruck habe, dass diese Erklärungen besser verstanden würden – eine Einschätzung, die auch eine weitere Lehrperson teilt. Bei vier Lehrpersonen konnte zudem beobachtet werden, dass sie die Klasse oder eine/n Schüler/in zu einer (spontanen) Selbsteinschätzung anregten. So fragte eine Lehrperson einen Schüler, nachdem in Einzelarbeit eine Sprechaktivität vorbereitet worden war, ob er nun fünf Minuten zum Thema sprechen könne. Der Schüler antwortete mit „clairement pas“. Mit dieser Frage wollte die Lehrperson – wie sie im Interview sagte – erreichen, dass die Lernenden erkennen würden, dass sie den Wortschatz noch besser lernen müssten.

In den Interviews wurde deutlich, dass es die Lehrpersonen aus Zeitgründen und aufgrund der grossen Klassen generell als schwierig erachten, Rückmeldungen zur Sprechkompetenz zu geben. Zudem befürchteten die Lehrpersonen, die Lernenden mit Korrekturen zu demotivieren. Korrekturen oder Rückmeldungen werden deshalb oft an die ganze Klasse adressiert und sind nicht individualisiert. Auch geben in den Interviews alle Lehrpersonen an, in bestimmten Situationen bewusst nicht zu korrigieren. Als häufigsten Grund dafür nennen sie eine ausreichende Verständlichkeit. Was genau darunter verstanden wird, dürfte situationsabhängig sein. Einige Lehrpersonen erwähnen, dass sie sich bei einem festgestellten

Fehler fragen würden, ob ein deutscher Muttersprachler eine fehlerhafte Aussage dennoch verstehen würde. Je nachdem korrigieren sie – oder eben nicht. Der zweithäufigste genannte Grund, nicht zu korrigieren, betrifft die Motivation. So meint LP07: „J’ai pas envie de les corriger pour tout en fait [...] je peux pas toujours les casser“. Die Motivation war in den Interviews ein wiederkehrendes Thema. Es sei schwierig, die Schüler/innen überhaupt zum Sprechen zu bringen. Als Gründe vermuten die Lehrpersonen fehlenden Wortschatz oder Angst vor Fehlern bzw. dem Exponiert-Sein. So wird darauf geachtet, nicht (zu) viel zu korrigieren, weil ansonsten demotivierende Effekte befürchtet werden. Zumeist betonen die Lehrpersonen, die Lernenden zu ermutigen, Deutsch zu sprechen. Dabei seien Fehler zweitrangig: „c’est plus important pour moi qu’il parle allemand que qu’il parle JUSTE“ (LP03). Auch Lob und Komplimente werden bewusst eingesetzt, mit dem Ziel, die Motivation der Lernenden beziehungsweise deren Selbstwertgefühl zu stärken. Im Unterricht konnten diesbezüglich vor allem kurze Kommentare wie „sehr gut“ beobachtet werden (zu weiteren Resultaten der ersten Projektphase vgl. Peyer et al., 2023).

Mit der Maxime des Motivationserhalts lässt sich wohl auch erklären, weshalb insbesondere der implizitere Korrekturtyp der *Recasts* von den Lehrpersonen angewendet wird. Ähnlich wie Lehrpersonen in anderen Studien (Kamiya, 2016; Roothoof, 2014) äusserten einige Lehrpersonen Bedenken, bei den Lernenden mit ihren Korrekturen negative affektive Reaktionen hervorzurufen. Da *Recasts* sich jedoch in Lyster & Saitos (2010) Metaanalyse von Unterrichtsstudien als weniger effektiv als *Prompts* erwiesen haben, wäre ein vermehrter Einsatz letzterer wünschenswert. Selbstverständlich bleibt es trotzdem wichtig, auf die Motivation der Schüler/innen und ihre affektiven Reaktionen auf korrektives Feedback achtzugeben. Wie auch Roothoof (2014) festhält, ist es erstaunlich, dass es bislang kaum Studien zu den affektiven Reaktionen der Lernenden auf Korrekturen gibt, obwohl die Lehrpersonen offensichtlich Vermutungen dazu anstellen, die wiederum ihr Korrekturverhalten beeinflussen.

4.2 Wie werden Paarprüfungen auf der Sek I durchgeführt?

Wie oben erwähnt, wurden in acht Klassen authentische Sprechtests gefilmt. Die mündlichen Prüfungen wurden entweder als Paarprüfung durchgeführt (4 Klassen), als Einzelprüfung (1 Klasse) oder sie bestanden aus zwei Teilen, einem Paar- und einem Einzelprüfungsteil (3 Klassen). Im Folgenden wird das Prüfer/innenverhalten der vier Lehrpersonen, die Paarprüfungen ($n=21$) durchführten, näher beleuchtet. Dieses ist bei Paarprüfungen auf den tieferen Niveaus (A1-A2), auf denen sich die meisten gefilmten Schüler/innen befanden, deutlich wichtiger als auf höheren Niveaus, da gemäss dem Gemeinsamen europäischen Referenzrahmen (GeR) Lernende auf Niveau A2 noch kaum in der Lage sind, ein Gespräch selbst in Gang zu halten (Europarat, 2001: 79). Gemäss Galaczi erfordern Paarprüfungen auf Niveau A2 deshalb noch regelmässig Unterstützung und Kontrolle durch die Prüfungsleitung (Galaczi, 2014: 557).

Prüfungsaktivitäten und Ablauf der Prüfungen

Die Prüfungsaufgaben haben die Lehrpersonen selbst entwickelt, zumeist zusammen mit anderen Lehrpersonen ihrer Schulen. Abbildung 2 gibt einen Überblick über Lehrpersonen, Klassenstufen und Art der gefilmten Prüfungen.

Lehrperson	Eva ⁵	Mia	Lea	Tom
Unterrichtserfahrung	4 Jahre	25 Jahre	7 Jahre	20 Jahre
Klasse	Realklasse 8. Klasse (10H)	Realklasse 8. Klasse (10H)	Sek-Klasse 8. Klasse (10H)	Realklasse 9. Klasse (11H)
Beurteilungsaktivität	5 W-Fragen zum Thema Mode stellen und beantworten		5 Fragen zum Thema Taschengeld stellen und anhand eines Steckbriefs beantworten	Gemeinsam eine Aktivität organisieren (Fest, Ausflug)
Zweck der Prüfung	Überprüfung der Lernziele des Kapitels			Vorbereitung auf kantonale Abschlussprüfung

Abbildung 2: Angaben zu Lehrpersonen, Klassen und Art der Prüfung

Eva, Mia und Lea führten Tests zur Überprüfung der Lernziele eines Lehrbuch-Kapitels durch (*Achievement-Test*), Tom einen Test zur Vorbereitung auf die kantonale Abschlussprüfung (*Proficiency-Test*). Die Beurteilungsaktivitäten (inklusive Instruktionen) aller vier Lehrpersonen waren in Deutsch verfasst. Eva und Mia erstellten ihre Beurteilungsaktivität gemeinsam, da sie an derselben Schule unterrichteten. In ihrem Test sollten die Schülerinnen und Schüler über Mode sprechen und jeweils mindestens fünf Fragen dazu stellen und beantworten. Als Hilfestellungen waren auf dem Aufgabenblatt verschiedene Interrogativpronomen und einige Verben auf Deutsch abgedruckt. In Leas Prüfung sollten die Schüler/innen über Taschengeld sprechen und fünf Fragen zum Thema stellen und beantworten. Dazu bekamen sie einen Steckbrief eines Jugendlichen, der u. a. Informationen (zumeist in Form von Bildern) über die Höhe des erhaltenen Taschengeldes, den Schülerjob und

5 Eigennamen (von Personen und in den Transkripten genannten Orten) wurden pseudonymisiert.

das Sparziel enthielt. Ebenfalls waren einige Redemittel wie „verdienen“, „arbeiten“ oder „genug/zu wenig“ auf Deutsch abgedruckt. In Toms Prüfung mussten die Schüler/innen zu zweit eine Aktivität organisieren (Ausflug ins Kino, ins Schwimmbad usw.) und verschiedene Punkte wie Zeit, Transportmittel usw. besprechen. Auch sein Aufgabenblatt enthielt einige Redemittel auf Deutsch. Alle Lehrpersonen gaben den Schüler/innen Vorbereitungszeit, während der diese sich jedoch nichts notieren durften. Die gefilmten Prüfungen dauerten zwischen drei und sechs Minuten.

Wann und wie intervenieren Lehrpersonen auf Sekundarstufe I bei mündlichen (Paar-) Prüfungen?

Die vier beobachteten Lehrpersonen verhalten sich bei den mündlichen Paarprüfungen sehr unterschiedlich. So zeigt sich sowohl bei der Häufigkeit der Interventionen (von sehr gering bei Tom bis häufig bei Lea) als auch bei der Vielfalt ihrer Funktionen eine grosse Variabilität zwischen den Lehrpersonen: Tom greift auch bei offensichtlichen Interaktionsproblemen nicht in eine Prüfung ein, sondern nur um die Schüler/innen zu erinnern, ihre Diskussion am Ende einer Prüfung zusammenzufassen. Die anderen drei Lehrpersonen hingegen intervenieren nicht nur zum Zweck des Prüfungsmanagements (z.B. Aufforderung, eine Frage zu stellen), sondern auch, um die Kandidat/innen bei Interaktionsproblemen (Verständnisschwierigkeiten, Wortsuche usw.) zu unterstützen und ihnen positives (meist nonverbales) Feedback zu geben. Mia und Lea greifen zudem manchmal auch in eine Prüfung ein, um die Äusserung eines Kandidaten/einer Kandidatin zu korrigieren, in der Regel durch eine Klärungsfrage oder einen *Recast*. Darüber hinaus greift Lea mehrmals in eine Prüfung ein, um zu überprüfen, ob die Schüler/innen Strukturen verwenden können, die die Klasse vor kurzem gelernt hat, wie z.B. kausale oder finite Nebensätze. Ein Grund, weshalb Lea während der Prüfungen immer wieder grammatische Strukturen zu elizitieren versucht, dürfte sein, dass es Freiburger Lehrpersonen seit einigen Jahren verboten ist, Grammatik- und Wortschatztests im Fremdsprachenunterricht durchzuführen. Begründet wird dies damit, dass Grammatik- und Wortschatzkenntnisse bei der Überprüfung der kommunikativen Sprachkompetenzen mitgetestet werden sollten (SEnOF, 2018).

Insgesamt dürfte die beträchtliche Variabilität im Verhalten der Lehrpersonen auf verschiedene Aspekte zurückzuführen sein, z.B. auf Unterschiede beim Prüfungszweck und den verwendeten Prüfungsaktivitäten. Es ist wahrscheinlich kein Zufall, dass nur Tom, der kaum in die Prüfungen eingreift, eine Beurteilungsaktivität verwendet, die zumindest in einigen Punkten einem Task entspricht (z.B. in Bezug auf Zielorientierung oder die Authentizität) und stark an Prüfungsaufgaben erinnert, die in standardisierten *Proficiency-Tests* für das Niveau A2 anzutreffen sind. Die Aktivitäten der anderen Lehrpersonen haben z.B. ein im Voraus geregeltes *Turn-Taking* (Sprecherwechsel) und erfordern weder einen natürlichen Sprachgebrauch noch fördern sie das gemeinsame Aushandeln – Aspekte, die in der Literatur als Vorteile von Paarprüfungen diskutiert werden (z.B. May, 2009; Rydell, 2019). Insbesondere Leas Aktivität weist zudem Mängel bei der Verständlichkeit auf (u. a. ein irrefüh-

rendes Foto auf dem Aufgabenblatt), so dass sie während der Prüfungen regelmässig Fragen zur Aufgabe beantworten muss. Darüber hinaus scheint auch die Persönlichkeit der Lehrpersonen eine wichtige Rolle dabei zu spielen, wie oft und auf welche Art und Weise sie in Paarprüfungen eingreifen, wie das unterschiedliche Verhalten von Eva und Mia zeigt, die dieselbe Aktivität mit Schüler/innen desselben Niveaus verwendet haben.

Welchen Effekt haben verschiedene Typen von Interventionen auf die Peer-Interaktion?

In diesem Abschnitt werden typische und wiederholt beobachtete Interventionsstrategien im Hinblick auf ihre Auswirkungen auf das Fortschreiten der Peer-Interaktion analysiert, beginnend mit denjenigen, die diese in Gang halten, und endend mit jenen, die sie verlangsamten.

Eine der wirksamsten Interventionen, um die Peer-Interaktion zu unterstützen, ist gemäss unseren Beobachtungen die Reformulierung einer schwer verständlichen Frage, eine Art Mediations-Strategie, die Mia regelmässig erfolgreich anwendet. Exzerpt 1 ist ein typisches Beispiel dafür.

01 S1: äm:: (6.1) wo (2.9) wo: (2.8) <non> (1.3) ((flüstert unv., ca. 1 Sek.))
 02 welche (1.3) trage ((klopft bei jedem Wort mit dem Zeigfinger auf den Tisch)) (-)
 03 schwarz (-) t-shirt ((schaut zu S2))
 04 (1.0)
 05 Mia: ((wendet den Blick zu S2)) trägst du schwarze t-shirts
 06 S2: äh ja ich: äh: trage (-) äh schwarze t-shirt

Exzerpt 1: Reformulierung einer schwer verständlichen Frage (Mia, Prüfung 5)

Stockend und unterbrochen von mehreren längeren Pausen versucht die Schülerin S1 in Zeile 1 eine Frage mit dem Fragepronomen „wo“ zu formulieren, verwirft dann aber die begonnene Frage wieder und beginnt in Zeile 2 eine weitere mit „welche“ zu bilden. Letztlich bleibt der Sinn der Frage jedoch unklar (welche trage schwarz T-Shirt?), sodass Mia eingreift und direkt an die andere Schülerin S2 gewandt folgende Frage stellt: trägst du schwarze T-Shirts? S2 kann diese Frage daraufhin ohne längeres Zögern beantworten. Offen bleibt, ob Mia die Frage von S1 korrekt, d.h. im Sinne der Schülerin, reformulierte und ob ihre Intervention tatsächlich notwendig war oder ob die Kandidatinnen nicht in der Lage gewesen wären, ihre (möglichen) Interaktionsschwierigkeiten selbst zu lösen, indem S2 z.B. um Klärung gebeten hätte.

Eine weitere Interventionsstrategie, die sich in den beobachteten Prüfungen immer als förderlich für das Fortschreiten des Gesprächs erwies, ist die Aufforderung der Lehrperson zu improvisieren bzw. vom Steckbrief abzuweichen. Lea wendet diese Strategie ab und zu an, wenn ein/e Kandidat/in nach Worten sucht oder zu verstehen gibt, dass er/sie Formulierungsschwierigkeiten hat. In Exzerpt 2 hat der Schüler S7 Schwierigkeiten zu formulieren, dass er – wie auf seinem Steckbrief ersichtlich – als Nebenjob putzt, sodass ihn Lea ermuntert, „einfach eine Information [zu] geben“ (Z.4).

01 S8: hm und in was jobbst du ((schaut zu S7))
 02 S7: ich jobbe in äh (1.9) äh: <ah (non) ça je sais pas dire> (2.7) ah (1.0)
 03 Lea: du musst nicht GENAU sagen was auf dem foto ist du kannst a:uch
 04 ((kreisende Handbewegung)) (---) einfach eine information geben
 05 S7: okay ich jobben in (1.5) äh ich <help>
 06 ((schaut zu Lea, kratzt sich am Kopf)) [mein jobben ist zu helfen] (-) äh
 07 Lea: [((schaut zu S7, nickt mehrmals))]
 08 S7: neuest (-) äh: (----) neuest schu:\ (-) sch\ sch: öh schule
 09 Lea: [((nickt)) hm_hm
 10 S7: öh in der <co> <of> fribourg (-) mit (-) den mathematik und französisch
 11 Lea: sehr gut (-) fantastisch

Exzerpt 2: Aufforderung zum Improvisieren (Lea, Prüfung 4)

Wie Exzerpt 2 zeigt, kann S7 nach der Aufforderung zum Improvisieren stockend und mit gelegentlichen Rückgriffen auf das Englische und Französische („help“ statt helfen in Z. 5 und Z.6; „of“ und „CO“ für *cycle d'orientation*, d.h. Sekundarschule in Z.10), aber ohne weitere Hilfe in einfachen Worten ausdrücken, dass er an einer Sekundarschule Nachhilfe in Mathe und Französisch gibt. Dabei reicht aus, dass die Lehrperson ermutigend nickt, während er nach Worten sucht. Eine Aufforderung zum Improvisieren ermöglicht es also, zu überprüfen, was Kandidat/innen mit sehr begrenzten sprachlichen Mitteln ausdrücken können und ihnen evtl. zu einem Erfolgserlebnis zu verhelfen. Exzerpt 2 verweist zusätzlich auf unklare Erwartungen bzgl. der Aufgabenerfüllung: Dadurch, dass Lea diesem Schüler anstelle der Wiedergabe von vorgegeben Inhalten eine freie Produktion erlaubt, vereinfacht sie die Testaufgabe. Diese Hilfestellung müsste dann in die Bewertung einfließen.

Interventionen, die manchmal für das Fortschreiten der Peer-Interaktion förderlich sind, sind Bitten um Klärung in der Schulsprache, eine Strategie, die Mia regelmässig einsetzt.

01 S3: ö:m (5.8) ä:m (1.2) wo ist (1.1) äh die shoppen
 02 (0.9)
 03 Mia: ((schaut zu S3))
 04 S3: ((Blickkontakt mit Mia)) äm (1.5)
 05 Mia: <hm qu'est-ce que t'aimerais dire [S3]>
 06 S3: [äh]
 07 S3: <où il va faire shop\ le shopping>
 08 Mia: oh wo shoppst du ((schaut zu S4)) [wo gehst du] shoppen
 09 S3: [ah wo shopps\
 10 S4: ä:m (2.3) ich shoppe ((schaut zu Mia, die auf Raster schaut, 1.0 Sek.))
 11 ((starrt ins Leere, 0.8 Sek.)) äh: ((atmet tief ein)) (-) in fribourg

Exzerpt 3: Klärungsfrage in der Schulsprache (Mia, Prüfung 4)

In Exzerpt 3 formuliert der Schüler S3 nach langem Zögern eine fehlerhafte und schwer verständliche Frage mit dem Fragepronomen „wo“ (Z.1). Mia schaut ihn daraufhin an (Z.3) und wechselt nach einer kurzen Pause ins Französische, um zu fragen, was er sagen wollte, eine Frage, die S3 ebenfalls auf Französisch beantwortet. Mia formuliert daraufhin eine korrekte

Frage auf Deutsch (Z.8), die S3 in Zeile 9 sofort zu wiederholen beginnt. Die Intervention von Mia führt also zu einem partiellen „learner uptake“ (Lyster & Ranta, 1997: 49). Sein Kollege S4 kann schliesslich zögerlich eine richtige Antwort geben, ohne selbst um Hilfe bitten zu müssen. Mia setzt Klärungsfragen regelmässig ein, obwohl sie nur in etwa der Hälfte der Fälle zu einer (schnellen) Lösung eines Interaktionsproblems führen. Auch reicht es – wie in Exzerpt 3 ersichtlich – nicht aus, den/die Schüler/in einfach in der Schulsprache zu fragen, was er/sie eigentlich sagen wollte. In einem weiteren Schritt muss die Lehrperson dann die Antwort ins Deutsche übersetzen. Diese Klärungsfragen deuten zudem darauf hin, dass die Schülerinnen und Schüler noch nicht über die notwendigen Strategien verfügen, um ihre Interaktionsschwierigkeiten selbst zu lösen, was von Lernenden auf dem Niveau A2 erwartet werden könnte (Council of Europe, 2018: 84). Generell scheinen die beobachteten Schülerinnen und Schüler wenig Strategien zur Verfügung zu haben, insbesondere wenig Kompensationsstrategien bei einer Wortsuche.

Einige Interventionen sind – unseren Beobachtungen zufolge – nicht förderlich für das Fortschreiten der Peer-Interaktion, insbesondere das Elizitieren von Wörtern und Strukturen. In Exzerpt 4 soll der Schüler S7 sagen, wofür er sein Taschengeld ausgibt und dazu die Informationen im Steckbrief verwenden, wo auf dem entsprechenden Bild ein Geschenk, Pommes frites und eine Wurst abgebildet sind.

01 S7: (--) öm (-) ich kaufen mein gelb für geschenke un öh geschenke für mein freundIN
 02 Lea: ((lächelt und nickt))
 03 S7: (-) und öh (1.7) äh: (1.2) <saucisse> ((hebt seinen Blick zu Lea))
 04 Lea: ((schaut S7 an, 1.0 Sek.)) was ist das wort auf deutsch für <saucisse> ((Geste mit Handflächen nach oben))
 05
 06 S7: ah <mais je [sais]>
 07 Lea: [oder kennst] du [öh]
 08 S7: [(<<'est>)] öh frül\ frül\ früst ((schaut Lea an))
 09 <non> (--) ((schaut auf Blatt))
 10 Lea: po
 11 S7: ((schaut Lea an))
 12 Lea: (---) pom
 13 S7: (-) pom f[rüst]
 14 Lea: [pommes]
 15 S7: pommes
 16 Lea: pommes ((nickt)) okay ((schaut S8 an))

Exzerpt 4: Elizitieren eines Wortes (Lea, Prüfung 4)

Da S7 sich nicht an das deutsche Wort für „Wurst“ erinnern kann, verwendet er das französische Wort „saucisse“ und schaut die Lehrerin hilfeschend an (Z.3). Lea fragt darauf, was denn das deutsche Wort für „saucisse“ sei (Z.4). Da S7 es aber eben nicht (gleich) nennen kann, versucht sie daraufhin, die ebenfalls auf dem Teller abgebildeten Pommes frites zu elizitieren, indem sie fragt „oder kennst du“ (Z.7). Allerdings scheint S7 immer noch auf der Suche nach dem Wort für „Wurst“ zu sein, denn er sagt darauf: „c'est öh frül\ frül\ früst“ (Z.9), was etwas an den Plural „Würste“ erinnert. Lea scheint es nicht gehört zu

haben, denn sie fährt fort, das Wort für „Pommes“ zu elizitieren, indem sie die erste Silbe des Wortes sagt. Letztlich braucht es drei Interventionen von Lea, bis S7 schliesslich in der Lage ist, das Wort „Pommes“ zu sagen bzw. nachzusprechen. Dieses Beispiel zeigt, dass das Elizitieren eines Wortes durch die Lehrperson im Falle einer Wortsuche nicht unbedingt hilfreich für das Fortschreiten der Interaktion ist. Im Gegenteil, die obige Intervention von Lea hat die Interaktion zwischen den beiden Kandidaten für eine längere Zeit unterbrochen und den Kandidaten S7 wohl auch verwirrt.

Häufiger noch als einzelne Wörter elizitiert Lea komplexere Strukturen, z. B. Nebensätze mit „damit“. Dies unterbricht die Peer-Interaktion immer für einen längeren Zeitraum, da mindestens zwei Interventionen und meist auch eine Korrektur durch die Lehrperson erforderlich sind. Lea geht es dabei wohl primär darum, zu überprüfen, ob die Kandidaten eine gerade im Unterricht behandelte Struktur verwenden können. Diese Intervention zielt also nicht darauf ab, die weitere Interaktion zwischen den beiden Kandidaten zu gewährleisten, und dient somit nicht einem Hauptziel von Paarprüfungen, nämlich der Überprüfung der Interaktionskompetenz.

4.3 Welchen Nutzen sehen Lehrpersonen in der summativen Beurteilung des Sprechens?

In der zweiten Forschungsphase wurden die sieben teilnehmenden Lehrpersonen in den retrospektiven Interviews gefragt, wie sie die durch die Prüfung gewonnenen Erkenntnisse nutzen würden („Qu'allez vous faire avec les informations que vous avez obtenues grâce à cette épreuve?“). Die meisten Lehrpersonen kamen bei dieser Frage zunächst darauf zu sprechen, welche Rückmeldung sie den Schüler/innen nach der Prüfung geben würden. Einige Lehrpersonen geben den Schüler/innen das ausgefüllte Raster inklusive Note (und allenfalls wenigen Kommentaren zu Fehlern) zurück, bezweifeln jedoch alle, dass die Schüler/innen dieses Feedback für das weitere Lernen nutzen würden. Eine erklärt deshalb allen auch noch mündlich, wie die Punktezahl bzw. die Note zu Stande gekommen ist. Weitere zwei Lehrpersonen, die ebenfalls daran zweifeln, dass die Schüler/innen ein schriftliches Feedback aufnehmen, geben jedem/r Schüler/in die Möglichkeit, gemeinsam mit ihnen die Audioaufnahme der Prüfung anzuhören, um dabei individuelles mündliches Feedback zu erhalten. Einige weitere Lehrpersonen geben den Schüler/innen ebenfalls keine schriftliche Rückmeldung, sondern stehen für Fragen (zur Note bzw. der individuellen Leistung) zur Verfügung. Zusätzlich zur individuellen Rückmeldung geben einige Lehrpersonen auch an, der ganzen Klasse ein kollektives Feedback zu häufig beobachteten Fehlern zu geben. Manche Lehrpersonen erwähnen zudem, bei der Rückmeldung zur Prüfung die Klasse oder besonders ängstliche Schüler/innen zu ermuntern bzw. zu versuchen, das Selbstbewusstsein der Schüler/innen zu stärken. Zwei Lehrpersonen merken zusätzlich an, die gewonnenen Informationen zu nutzen, um die Prüfung selbst bzw. das Beurteilungsraster zu überarbeiten. Eine weitere entschied sich während der Prüfung spontan, die Prüfung nicht zu benoten, sondern sie nochmals durchzuführen und die Schüler/innen

noch besser darauf vorzubereiten.

Im Weiteren wurden die sieben Lehrpersonen auch gefragt, ob ihrer Meinung nach Sprechprüfungen für die Schüler/innen nützlich sind, um Fortschritte in Deutsch zu machen. Die Lehrpersonen waren sich bei dieser Frage uneins, die Einstellungen waren je etwa zur Hälfte negativ bzw. positiv. So sieht Lea generell keinen Nutzen in mündlichen Prüfungen: „c'est plus une source de stress ces évaluations que une source de progression“. Zum Lernen motivieren könne man die Schüler/innen viel eher, indem man ein Treffen mit einer Partnerklasse aus der Deutschschweiz organisiere. Auch eine andere Lehrperson äussert im Interview Zweifel daran, ob ihre Schüler/innen von mündlichen Prüfungen profitieren würden, da sie in diesem Alter den Sinn noch nicht sehen würden, Deutsch zu lernen. Zwei weitere Lehrpersonen differenzieren zwischen dem niedrigeren und dem höheren Anforderungsniveau: Den schwächeren Klassen, mit denen sie an der Studie teilgenommen haben, würden mündliche Prüfungen nichts bringen, da diese Schüler/innen sich sowieso nicht auf die Prüfung vorbereiten würden, während Klassen mit höherem Anforderungsniveau auf eine Prüfung lernen würden. Die restlichen drei Lehrpersonen sind mündlichen Prüfungen gegenüber positiver eingestellt. Eva und Tom, die beide mit einer Klasse mit niedrigem Anforderungsniveau an der Studie teilgenommen haben, sehen die Prüfungen primär als ein (Druck-)Mittel, die Schüler/innen zum „motivierteren Arbeiten“ bzw. „seriöseren Üben“ zu bewegen. Anders argumentiert die dritte Lehrperson: Sie glaubt, dass mündliche Prüfungen von allen Prüfungen am meisten Fortschritt bewirken: Zum einen würden sich die Schüler/innen ihres Niveaus bewusst bzw. ob sie bereits in einer Alltagssituation interagieren könnten oder nicht. So würden sich die Schüler/innen bestärkt fühlen oder merken, dass sie sich mehr Mühe geben müssten. Gleichzeitig schätzt diese Lehrperson mündliche Prüfungen, weil sie den Schüler/innen so direkter helfen könne als bei schriftlichen Prüfungen und ihnen direkt anschliessend ein mündliches Feedback geben könne.

Insgesamt lässt sich bei den interviewten Lehrpersonen eine gewisse Resignation feststellen, was den Nutzen von mündlichen Prüfungen und insbesondere die Nachhaltigkeit ihrer Rückmeldungen auf diese Prüfungen anbelangt, bzw. eine gewisse Ratlosigkeit, wie sie es besser machen könnten. Sie zweifeln mehrheitlich daran, dass schriftliches Feedback (z. B. das ausgefüllte Raster) von den Schüler/innen aufgenommen bzw. für das weitere Lernen genutzt wird. Aus der Forschung ist bekannt, dass Feedback dann besonders lernförderlich ist, wenn es zeitnah erfolgt, konkret und beschreibend ist sowie aufzeigt, was und wie die Lernenden (sich) verbessern können (u. a. Luoma, 2004; Hattie & Timperley, 2007). Ein solches konkretes, zeitlich unmittelbar an die Prüfung anschliessendes (mündliches) Feedback, konnten wir – vermutlich aus zeitlichen und praktischen Gründen – nur in Ansätzen bei zwei Lehrpersonen beobachten. Es erstaunt deshalb nicht, dass ca. die Hälfte der sieben interviewten Lehrpersonen den Nutzen von Sprechprüfungen aus unterschiedlichen Gründen gerade bei Klassen mit niedrigerem Anforderungsniveau eher kritisch sehen.

4.4 Welche Bedürfnisse haben die Lehrpersonen bzgl. der summativen Beurteilung des Sprechens?

In den Interviews mit den Lehrpersonen wurde deutlich, dass die Durchführung von Sprechprüfungen mit verschiedenen Schwierigkeiten verbunden ist: So sind sie zeitaufwändig, besonders auch dann, wenn die Lehrpersonen die Sprechleistungen aufnehmen und die Beurteilung erst anschliessend durchführen. Wiederholt wurden aber auch disziplinarische Schwierigkeiten erwähnt, die es den Lehrpersonen verunmöglichen, die Prüfungen ausserhalb des Klassenzimmers durchzuführen, da die Klasse nicht unbeaufsichtigt gelassen werden könne. Vor diesem Hintergrund wurden die zwölf Lehrpersonen der ersten Phase gefragt, ob sie sich Hilfestellungen oder weitere Materialien für die summative Beurteilung des Sprechens wünschen würden. Bei dieser Frage unterscheiden sich die Antworten der Freiburger und Neuenburger Lehrpersonen deutlich, was auf unterschiedliche Rahmenbedingungen zurückzuführen sein dürfte. Im Kanton Freiburg ist die Durchführung von mehreren mündlichen Prüfungen pro Jahr (in Abhängigkeit der Anzahl Wochenstunden) verpflichtend, ebenso wie die Benutzung eines bestimmten Beurteilungsrasters, das zwischen vier Aspekten (Aufgabe, kommunikative Gestaltung, formale Richtigkeit und Aussprache/Intonation) unterscheidet. Der Kanton Neuenburg hingegen empfiehlt seinen Lehrpersonen lediglich, mündliche Prüfungen durchzuführen, macht jedoch keine Angaben zur Häufigkeit oder zur Art des zu benutzenden Beurteilungsrasters. Die vier Lehrpersonen aus Neuenburg erwähnen bei der Frage nach wünschenswerten Hilfestellungen verschiedentlich, dass ihnen die Pädagogische Hochschule nützliches Material zur Beurteilung zur Verfügung stelle. Weiteres, konkretes, direkt einsetzbares Material zur Beurteilung des Sprechens bzw. auch für verschiedene Gruppen adaptierbares Material wäre aber willkommen.

Die acht Freiburger Lehrpersonen äussern bei der Frage nach Hilfestellungen für die Beurteilung alle spontan ihre Meinung zum obligatorischen Raster. Die Hälfte von ihnen äussert deutliche Kritik daran, während zwei weitere Lehrpersonen Teilaspekte kritisieren und die restlichen zwei mit dem Instrument zufrieden sind. Die kritischen Lehrpersonen sind sich einig, dass das Raster nicht genügend differenziere und verfeinert werden müsste. Einige erwähnen, dass die Interpretation der Kategorien schwierig sei: „erfüllt/nicht erfüllt beim Raster – was heisst das konkret?“ (Eva), bzw. subjektiv sei: „Des fois j'ai l'impression que c'est une appréciation très personnelle finalement“ (LP06), bzw. „On évalue théoriquement tous de la même façon. Voilà je dis bien théoriquement parce que je suis pas tout à fait convaincue de ça“ (LP04). Einige äussern auch grundsätzlichere Kritik, wie Lea, die findet, dass die Beurteilung mit dem Raster den Schüler/innen nicht gerecht werde und sie damit zu defizitorientiert bewerten würde:

J'aime pas vraiment travailler avec ces grilles cantonales. Je trouve que je considère pas assez l'élève et que je me concentre sur ces quatre critères et que chaque fois qu'il y a un truc qui coince je me dis ah ben là j'enlève un point. Puis c'est vraiment j'ai l'impression une évaluation vers le bas plutôt que de valoriser

ce qu'ils ont déjà su [...]. J'ai pas l'impression que c'est le meilleur outil qu'on ait pour évaluer l'oralité.

Die interviewten Freiburger Lehrpersonen würden sich zumeist wünschen, dass das Raster konkretisiert und verfeinert würde, es dazu Anschauungsbeispiele zur Beurteilung geben würde bzw. dass es nicht mehr obligatorisch wäre. Ansonsten wünschen sie sich – wenn überhaupt – eher weiteres, konkretes und differenzierendes Material, um das Sprechen auf vielseitige Weise zu *üben*.

Die Kritik diverser Lehrpersonen am Freiburger Raster ist unseres Erachtens begründet, da das Raster in verschiedener Hinsicht zu unspezifisch ist sowie auch anderweitige Mängel aufweist. So lassen z.B. die Beschreibungen der Kriterien viel Interpretationsspielraum zu, da das Raster keinen Bezug zu Kompetenzbeschreibungen des GeR oder des Lehrplans herstellt und es keine Referenzbeispiele gibt (d.h. Aufnahmen von Schüler/innen, die kommentiert und eingestuft sind). So sind z.B. die Abstufungen für das Kriterium „Aussprache und Intonation“ („grösstenteils angemessen, ausreichend angemessen, nur teilweise angemessen, kaum angemessen, nicht angemessen“) ohne Referenzbeispiele eigentlich nur willkürlich anwendbar. Insofern erstaunt es nicht, dass drei Lehrpersonen in den Interviews spontan angeben, für das Kriterium „Aussprache und Intonation“ allen Schülern und Schülerinnen die gleiche (genügende) Anzahl Punkte zu geben, da sie Mühe hätten, sie diesbezüglich einzuschätzen. Ihre Unsicherheit begründen sie zum einen mit der Vagheit des Rasters, zum andern aber auch damit, dass die Schüler/innen schlicht zu wenig sprechen würden, um ihre Leistung beurteilen zu können. Ein weiterer Schwachpunkt des Rasters ist die Ansammlung verschiedenster Aspekte unter dem Kriterium „Kommunikative Gestaltung“, das unter anderem die „Variation der Satzstruktur“, „die Variation des Wortschatzes“, die „Verwirklichung der Sprechabsicht“ inklusive „korrektem Wortgebrauch“ sowie die „Flüssigkeit der Rede“ umfasst, Aspekte, die z.B. im Beurteilungsraster zum Sprechen von Lingualevel auf die Kriterien „Spektrum“, „Korrektheit“ und „Flüssigkeit“ aufgeteilt sind (Lenz & Studer, 2007). Zur Kritik am Raster selbst kommt hinzu, dass verschiedene Lehrpersonen angemerkt haben, nicht oder nur ungenügend im Umgang mit dem Raster geschult worden zu sein, was die Reliabilität des Instruments zusätzlich in Frage stellt.

4.5 Was denken die Schüler/innen über Sprechprüfungen?

Im Rahmen einer schriftlichen Befragung ($n=254$) und retrospektiver Interviews ($n=84$) äuserten auch die Schüler/innen ihre Ansichten und Wünsche zu Sprechprüfungen. Die Wichtigsten sollen hier vorgestellt werden (für mehr Informationen zur Sicht der Schüler/innen vgl. Lüthi et al., eingereicht).

Im Fragebogen gaben die Schüler/innen u. a. an, wie nützlich sie verschiedene Prüfungsformate finden, um ihr Deutsch zu verbessern (Abb. 3). Es zeigt sich, dass die Schüler/innen verschiedene Formate nützlich finden, wobei die Paarprüfungen mit 80% (eher) zustimmenden Schüler/innen als am nützlichsten erachtet werden, hilfreicher als Einzelprü-

fungen und Vorträge. In den Interviews wurden die Schüler/innen zudem gefragt, ob ihnen Sprechprüfungen helfen, Fortschritte im Deutschen zu erzielen. Dabei fanden 57% ($n=48$) bzw. 23% ($n=19$) der Schüler/innen die Prüfungen nützlich oder zumindest teilweise nützlich, während 18% der Schüler/innen ($n=15$) sie nicht als nützlich einschätzten. Die Schüler/innen begründeten ihre positive Einstellung hauptsächlich damit, dass sie mit Sprechprüfungen die Sprechkompetenz aufbauen könnten (14 SuS, 17%), dass sie später im Leben (z.B. Lehre, Arbeit) oder bei Reisen in deutschsprachigen Gebieten Deutsch sprechen müssten (10 SuS, 12%), oder dass sie dadurch die Aussprache üben könnten (9 SuS, 10%). Viele der Argumente laufen also darauf hinaus, dass Sprechtests eine Gelegenheit bieten, das Sprechen oder Aspekte der Sprechkompetenz zu üben bzw. dass Sprechen eine wichtige und nützliche Fähigkeit ist und daher Sprechtests nützlich sind. Die Schüler/innen, die Sprechprüfungen nutzlos fanden, gaben verschiedene Gründe an, der häufigste war jedoch mit Stress verbunden (6 SuS, 7%).

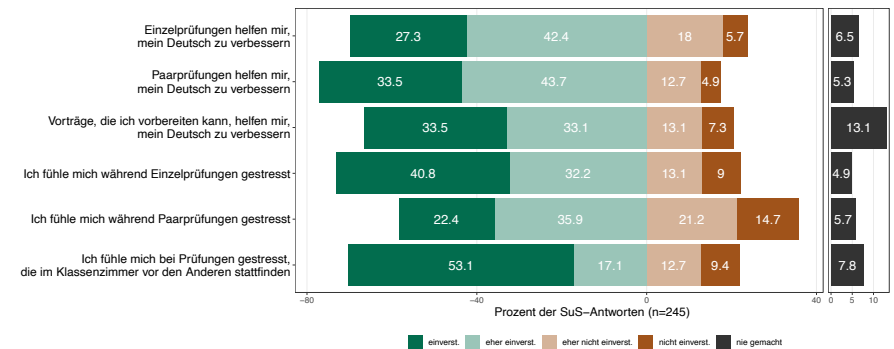


Abbildung 3: Resultate des SuS-Fragebogens zur Nützlichkeit und dem Stressgefühl bei verschiedenen Prüfungsformaten

Ebenfalls sollten die Schüler/innen im Fragebogen angeben, ob sie bei verschiedenen mündlichen Prüfungsformaten gestresst sind (Abb. 3). Auffallend ist, dass die Schüler/innen verschiedene Prüfungsformate, die sie zwar als nützlich erachten, dennoch als stressig einschätzen, wobei die Paarprüfung sowohl als das am wenigsten stressige und als auch als das nützlichste Format angesehen wird. Während Einzelprüfungen von drei Vierteln der Schüler/innen als stressig oder eher stressig empfunden werden, liegt der Anteil noch bei ca. 60%, wenn die Prüfung zu zweit abgelegt werden kann. Betrachtet man bei den Items „Ich fühle mich bei Einzel-/Paarprüfungen gestresst“ nur den Anteil der deutlich Zustimmenden (bzw. Gestressten), zeigt sich ein noch deutlicherer Unterschied: Während bei Einzelprüfungen knapp 41% der Schüler/innen klar gestresst ist, halbiert sich dieser Anteil bei Paarprüfungen auf 22%. Auch fühlen sich viele Schüler/innen deutlich gestresst, wenn sie die Prüfung exponiert vor der ganzen Klasse ablegen müssen. Darauf deutet der hohe Anteil an klarer Zustimmung (53%) beim Item „ich fühle mich bei Prüfungen gestresst,

die im Klassenzimmer vor den Anderen stattfinden“ hin.

Im Fragebogen wurden die Schüler/innen des Weiteren nach ihren Vorlieben bezüglich Rückmeldungen auf Sprechprüfungen befragt. Sie zeigten sich dabei an allen vorgeschlagenen Feedbacktypen interessiert. Die Reihenfolge der Zustimmung (von der höchsten zur geringsten) war 1) Ratschläge für die Verbesserung, 2) ein Kommentar, was in der Prüfung gut und weniger gut gemacht wurde, 3) eine Note, 4) ein ausgefülltes Beurteilungsraster, 5) ein Feedback zur Leistung der gesamten Klasse. Die Wünsche der Schülerinnen entsprechen damit nicht dem Feedback, das sie von ihren Lehrpersonen erhalten. Letztere geben immer eine Note und teilweise das ausgefüllte Beurteilungsraster oder ein Feedback an die ganze Klasse (siehe Kapitel 4.3). Das Feedback, für das sich die Schüler/innen am meisten interessieren – Ratschläge für die Verbesserung bzw. einen Kommentar zu was gut und weniger gut gemacht wurde – wird von den befragten Lehrpersonen nicht bzw. nur sehr selten gegeben.

Im Fragebogen gaben die Schüler/innen im Weiteren auch kurz Auskunft, was sie mit dem erhaltenen Feedback machen würden. Bei den Antworten auf diese Frage ist zu beachten, dass nicht alle Schüler/innen gleiches oder gleich viel Feedback erhalten. Ihre Antworten lassen sich drei Kategorien zuordnen:⁶ Erstens gibt es Schüler/innen, die angeben, das Feedback zur Verbesserung zu nutzen ($n=121$, 57%). Dabei machen einige sehr vage Aussagen wie „Je m’améliore“; andere sagen zwar, ihre Fehler oder Schwächen zu verbessern, geben aber keinen Hinweis auf die Arbeit, die sie dafür leisten; während wieder andere auf einen Aufwand verweisen, den sie betreiben, z. B. „Guardo dove devo migliorare di più e mi alleno a casa e provo a migliorare“. Dazu kommen einige Schüler/innen, die das Feedback (bei der Vorbereitung) für die nächste Prüfung nutzen, wobei auch hier in den allermeisten Fällen unklar bleibt, woran oder wie gearbeitet wird. Zweitens gibt es Schüler/innen, die angeben, sich das Feedback anzuschauen ($n=66$, 30%): Einige konzentrieren sich dabei auf Fehler oder Schwächen, während andere nicht genauer beschrieben, worauf sie achten. Drittens gibt es Schüler/innen, die das Feedback nicht weiter nutzen ($n=47$, 19%). Einige beantworteten die Frage mit „nichts“, andere nur damit, dass sie das Feedback abheften oder ihren Eltern zeigen würden. Die Antworten der Schüler/innen weisen darauf hin, dass vielen das Potential von Feedback für das weitere Lernen bewusst ist. Die oftmals sehr vagen Aussagen lassen aber auch vermuten, dass die Schüler/innen das Feedback kaum aufnehmen und für das weitere Lernen nutzen (können), wobei sich auch die Frage stellt, inwiefern das Feedback, das die Lehrpersonen geben, für die Schüler/innen aussagekräftig ist (vgl. Kap. 4.3).

6 Einzelne Antworten wurden sowohl der ersten Gruppe („für weiteres Lernen nutzen“) als auch der zweiten Gruppe („Feedback anschauen“) zugeordnet.

5 Fazit

Unser Forschungsprojekt hatte zum Ziel, sowohl formative als auch summative Praktiken zur Beurteilung der Sprechkompetenz im DaF-Unterricht der Sekundarstufe I zu dokumentieren. Zudem sollten die diesbezüglichen Ansichten der Lehrpersonen und bzgl. summativem Assessment auch diejenigen der Schüler/innen erfasst werden. Während der in der ersten Projektphase beobachteten DaF-Lektionen konnten vor allem spontane Korrekturen beobachtet werden, jedoch kaum geplante formative Beurteilungsanlässe (vgl. Davison & Leung, 2009, für eine Übersicht über Typen der formativen Beurteilung). Betrachtet man die Beurteilungspraktiken vor dem Hintergrund, dass Beurteilungsanlässe im Klassenzimmer das Ziel verfolgen sollten, das Lernen der Schüler/innen zu verbessern, so fällt zudem auf, dass wir während der Unterrichtsbeobachtungen (sowie der summativen Prüfungen) kaum je individuelles, lernförderliches Feedback zur Sprechkompetenz bzw. das Aufzeigen von Verbesserungsmöglichkeiten beobachten konnten, wobei angemerkt werden muss, dass wir letztlich bei jeder Lehrperson nur einen kurzen Einblick (von zwei Lektionen) in den Unterrichtsalltag gewinnen konnten. Schülerinnen und Schüler würden sich jedoch laut Fragebogen als Feedback besonders Ratschläge für die Verbesserung ihrer Sprechkompetenz wünschen (vgl. Kap. 4.5).

Wie Sprechprüfungen auf der Sekundarstufe I durchgeführt werden und Lehrpersonen mit dieser komplexen Aufgabe umgehen, sollte in der zweiten Projektphase erforscht werden. Bei der Durchsicht der gefilmten authentischen Sprechprüfungen aus acht Klassen zeigte sich besonders bei der Durchführung von Paarprüfungen eine grosse Variabilität im Lehrpersonenverhalten. Während eine Lehrperson (Tom) so gut wie nie in eine Prüfung eingreift, auch nicht bei offensichtlichen Interaktionsproblemen der Kandidat/innen, erteilt eine andere Lehrperson (Lea) während der Prüfung schon fast Sprachunterricht und greift regelmässig und zu unterschiedlichen Zwecken in die Paarprüfungen ein: für das Prüfungsmanagement, bei Interaktionsproblemen, um zu korrigieren, um positives Feedback zu geben oder auch, um zu überprüfen, ob gerade unterrichtete Strukturen schon aktiv angewendet werden können. Eine Konversationsanalyse der Paarprüfungen konnte zeigen, dass sich nicht alle Interventionen positiv auf die Interaktion der Kandidat/innen auswirken. Während sich Aufforderungen zum Improvisieren oder die Reformulierung einer schwer verständlichen Frage für den Fortgang der Peer-Interaktion als förderlich erweisen, führt insbesondere das Elizitieren von Wörtern und Strukturen zu längeren Unterbrechungen der Peer-Interaktion. Das Elizitieren steht somit im Widerspruch zum eigentlichen Ziel von Paarprüfungen, der Überprüfung der mündlichen Interaktion. Auch führt es dazu, dass die in der Fachliteratur genannten Vorteile von Paarprüfungen, etwa dass Kandidat/innen mehr Verantwortung übernehmen würden und weniger reaktiv seien als bei Einzelprüfungen (Taylor & Wigglesworth, 2009; Rydell, 2019), nicht zum Tragen kommen können. Das unterschiedliche Prüferverhalten der vier untersuchten Lehrpersonen dürfte verschiedene Gründe haben, wie z. B. ein unterschiedlicher Prüfungszweck. So führt Tom einen vom Lehr-

werk unabhängigen *Proficiency*-Test zur Überprüfung der mündlichen Interaktionskompetenz durch, während die anderen drei Lehrpersonen die Lernziele des Lehrbuchkapitels überprüfen möchten, zu denen auch der Erwerb bestimmter Wörter und Strukturen gehört. Das Überprüfen der Interaktionskompetenz stand also bei diesen *Achievement*-Tests wohl nicht im Vordergrund, was allerdings in sich widersprüchlich ist, weil es sich ja trotzdem um eine Sprechprüfung handelt. Allgemein zeigt sich, dass Lehrpersonen mit Paarprüfungen unterschiedliche und auch divergierende Ziele verfolgen: So hat die Elizitierung von Strukturen nicht mehr viel mit der Überprüfung von Interaktionskompetenz zu tun.

Letztlich dürfte das unterschiedliche Prüferverhalten auch damit zu tun haben, dass Lehrpersonen – anders als bei standardisierten Tests – während einer Prüfung mehrere Rollen innehaben: So sind sie sowohl Testadministrator/innen als auch Beurteilende und gleichzeitig Lehrperson der Kandidat/innen sowie je nach dem auch Gesprächspartner/innen (Teasdale & Leung, 2000; Koizumi, 2022). Unsere Analyse zeigt, dass die vier Lehrpersonen ihre Rollen unterschiedlich gewichten: Während Tom hauptsächlich in der Rolle des Beurteilers und Testadministrators ist, ist Lea oft in der Rolle der Lehrperson, die Strukturen elizitiert und ab und zu korrigiert. Da Lea, Eva und Mia Prüfungsaktivitäten verwenden, die nicht den Qualitätskriterien von kommunikativen Tasks entsprechen, müssen sie auch deutlich mehr Prüfungsmanagement betreiben (z.B. Organisation des Sprecherwechsels). Unsere Ergebnisse zeigen daher auch, dass Lehrpersonen für ihre verschiedenen Rollen während einer Prüfung sensibilisiert und darauf hingewiesen werden sollten, dass diese Rollen nicht alle gleichzeitig und gleich gut erfüllt werden können und sollen. Während bei Einzelprüfungen die Interlokutor/innen-Rolle (und nicht gleichzeitig jene des/der Bewertenden) im Vordergrund steht, sollte bei Paarprüfungen der Fokus auf der Prüfungsleitung (und ggf. dem Bewerten) liegen.

Die Beurteilung der Leistungen stand in unserer Studie nicht im Fokus, jedoch wurde bei den Befragungen der Lehrpersonen deutlich, dass diese die von ihnen benutzten Beurteilungskriterien durchaus kritisch hinterfragen, jedoch – im Falle von Freiburg – an die Verwendung eines bestimmten Beurteilungsrasters gebunden sind. Dabei handelt es sich um ein *one-size-fits-all*-Instrument, das sowohl für Paar- als auch für Einzelprüfungen (z.B. Vorträge) eingesetzt werden soll, und sicherlich zu Recht als u. a. zu wenig konkret und differenziert kritisiert wird. Berechtigt scheint weiter auch die Kritik, dass Bewertungen mit dem Raster subjektiv ausfallen: Da dem Raster eine Anbindung an die GER-Niveaus fehlt und keine Referenzbeispiele vorliegen, folgen die Bewertungen zwangsläufig einem Erfahrungswert bzw. einer subjektiven Norm.

Insgesamt lässt sich festhalten, dass die Passung von Lernzielen, Prüfungszweck, Prüfungsaufgaben, Verhalten der Lehrperson und Beurteilungskriterien bei den von uns beobachteten Prüfungen im Hinblick auf eine möglichst faire Beurteilung noch verbessert werden sollte. Gleichzeitig sind die Lehrpersonen aber auch an Rahmenbedingungen (wie knappe zeitliche und personelle Ressourcen und je nach Kanton Vorgaben und Verbote) gebunden, die ihre komplexe Aufgabe nicht unbedingt vereinfachen. Um die Lehrpersonen in Bezug auf die Beurteilung der Sprechkompetenz bestmöglich zu unterstützen, erscheint es uns

wichtig, dass gerade die Lehrpersonen im Kanton Freiburg mehr Freiheiten erhalten, um selbst entscheiden zu können, wie die Lernziele am besten überprüft werden. Dazu gehört die Wahl eines auf Art und Zweck der Prüfung abgestimmten Beurteilungsrasters ebenso wie die Möglichkeit, gelegentlich Wortschatz- und Grammatiktests durchzuführen. So könnte verhindert werden, dass Lehrpersonen während einer mündlichen Prüfung ein interventionistisches Verhalten zeigen, nur um Strukturen zu elizitieren, deren Anwendung zur Bewältigung der Aufgabe kommunikativ gar nicht notwendig sind. Laut eigenen Aussagen mangelt es den Lehrpersonen nicht an Prüfungsmaterialien, hingegen zeigt unsere Studie in verschiedenen Bereichen Aus- und Weiterbildungs-Bedarf auf: insbesondere bei der Unterscheidung von Prüfungszwecken, der Aufgabenentwicklung, aber auch beim Erteilen von lernförderlichen Rückmeldungen, sowie beim Prüfer/innenverhalten. Um den Lehrpersonen die Auswirkungen unterschiedlichen Prüfer/innenverhaltens bewusst zu machen, können auch unsere aufbereiteten, im Projekt entstandenen Prüfungsvideos helfen, die auf Anfrage gerne zur Verfügung gestellt werden.

Einige gefilmte mündliche Prüfungen wurden für Aus- und Weiterbildungszwecke aufbereitet.

Personen, die an einer Hochschule arbeiten, können diese bei Moritz Sommet (moritz.sommet@unifr.ch), dem Leiter des Forschungsdatenarchivs des Kompetenzzentrums für Mehrsprachigkeit beziehen. Das dazu gehörige Begleitdokument (mit Transkripten, Informationen zu den Prüfungsaufgaben, sowie Expert/innen-Rating der Schüler/innen-Leistungen) findet sich unter:
<https://centre-plurilinguisme.ch/de/forschung/beurteilung-sprechkompetenz>.

Évaluation de la production orale en classe d'allemand langue étrangère:

pratiques et points de vue des enseignant-e-s et des élèves du
secondaire I

—

Elisabeth Peyer, Gabriela Lüthi, Nadia Ravazzini

1 Situation initiale

L'évaluation des compétences orales, qu'il s'agisse de réagir aux questions des élèves, de commenter des présentations ou de faire passer des examens, fait partie du quotidien de l'enseignant-e et joue un rôle central dans le processus d'enseignement et d'apprentissage. Elle permet aux élèves et à l'enseignant-e de faire le point sur ce qui est acquis et sur les lacunes qui subsistent. Idéalement, elle éclaire aussi l'enseignant-e sur la manière de mieux aider les élèves dans leurs apprentissages. Divers sondages ont toutefois révélé que les enseignant-e-s de langues étrangères tendent à manquer d'assurance en matière d'évaluation et souhaitent disposer de plus de connaissances et de savoirs pratiques sur différentes questions relatives à l'évaluation, et en particulier sur l'évaluation de l'expression orale (entre autres Berry et al., 2019; Vogt, 2018). Dans cette optique, la présente étude s'intéresse à ce que l'on appelle la littératie en évaluation des compétences linguistiques (*language assessment literacy*) des enseignant-e-s en matière de compétences de production orale. Dans le contexte étudié – l'enseignement de l'allemand comme langue étrangère au niveau secondaire I en Suisse romande et italienne –, la *language assessment literacy* revêt une importance particulière, car l'allemand est une discipline obligatoire et la note finale d'allemand peut être déterminante pour la suite du parcours scolaire des jeunes. Par *language assessment literacy*, on entend non seulement les connaissances et capacités requises pour concevoir et réaliser des sessions d'évaluations et pour évaluer les compétences, mais aussi la conscience de l'impact que ces évaluations peuvent avoir sur les individus et la société dans son ensemble (entre autres Inbar-Lourie, 2008; Fulcher, 2012; Vogt & Tsagari, 2014). Inbar-Lourie, par exemple, souligne que la (*language*) *assessment literacy* comprend également la faculté de poser des questions critiques sur les objectifs de l'évaluation, sur l'adéquation de l'outil utilisé, sur les conditions du test, ainsi que sur ce qui se passera en fonction des résultats obtenus (2008: 389).

L'évaluation des compétences orales se heurte à plusieurs obstacles. Elle s'avère difficile en raison du caractère éphémère de l'oralité notamment, et parce que différents facteurs, qui ne devraient pas avoir d'impact sur l'évaluation des compétences, peuvent intervenir dans le processus d'évaluation (variance non pertinente au construit mesuré) comme les tâches, les grilles critériées, les évaluateur-trice-s ou même les interlocuteur-trice-s (Isaacs, 2016; De Jong, 2023). Selon Csépes & Fekete, les enseignant-e-s doivent donc avoir une bonne compréhension des principes fondamentaux du développement des tâches, du processus d'évaluation et des grilles de critères pour évaluer de manière valable la production orale (2018: 99). De plus, une évaluation fondée sur des critères à l'aide d'une grille requiert une phase d'entraînement pour les enseignant-e-s et des exemples illustratifs (ibid.). Un autre défi de l'évaluation sommative de la production orale en classe réside dans le fait qu'elle exige beaucoup de temps et constitue une tâche particulièrement complexe pour les enseignant-e-s. En effet, ces dernier-ère-s doivent assumer simultanément plusieurs rôles qui, par exemple dans le cadre d'examens internationaux standardisés, sont

normalement attribués à différentes personnes : gestionnaire de test, évaluateur-trice et, selon le format du test, interlocuteur-trice (Teasdale & Leung, 2000; Koizumi, 2022). Sans oublier qu'en même temps, ils et elles restent aussi les enseignant-e-s des élèves concerné-e-s.

2 Questions de recherche

Dans le contexte décrit ci-dessus, le projet *Évaluation de la production orale en classe d'allemand langue étrangère* a examiné quelles pratiques d'évaluation formative et sommative les enseignant-e-s d'allemand langue étrangère (*Deutsch als Fremdsprache*, DaF) au niveau secondaire I en Suisse utilisent pour évaluer la production orale de leurs élèves, ainsi que leurs opinions à ce sujet. Alors que plusieurs sondages ont déjà été réalisés auprès d'enseignant-e-s sur leurs pratiques d'évaluation, leur formation et leurs éventuels besoins de formation continue en matière d'évaluation (entre autres Berry et al., 2019; Fulcher, 2012; Tsagari & Vogt, 2017; Vogt & Tsagari, 2014; Vogt, 2018), il n'existe encore que peu d'études – notamment dans le contexte de l'école obligatoire – qui observent les pratiques d'évaluation et les mettent en relation avec les points de vue exprimés par les enseignant-e-s. Ce constat surprend d'autant plus que les pratiques d'évaluation ne peuvent pas être relevées directement par des interviews ou des questionnaires ni considérées indépendamment de l'expérience et de l'attitude des enseignant-e-s (Berry et al., 2019; Scarino, 2013). La recherche ne s'étant en outre guère intéressée aux opinions et souhaits des apprenant-e-s directement concerné-e-s par les questions d'évaluation jusqu'à présent, nous avons inclus cet aspect dans notre étude, dont le but consiste en définitive à identifier la meilleure façon d'aider les enseignant-e-s comme les élèves à évaluer les compétences orales.

Le projet aborde les questions de recherche suivantes :

1. Quelles sont les pratiques d'évaluation formative des compétences orales utilisées par les enseignant-e-s du secondaire I et comment ces pratiques sont-elles explicitées ?
2. Pour les enseignant-e-s, quelle est l'utilité de l'évaluation sommative de l'expression orale et quels sont leurs besoins à cet égard en termes de soutien et de matériel ?
3. Comment les évaluations sommatives de la production orale se déroulent-elles ?
 - a. Lors des épreuves orales par deux, quand et comment les interventions des enseignant-e-s se déroulent-elles ?
 - b. Quels sont les effets des interventions des enseignant-e-s sur le déroulement ultérieur de l'interaction entre pairs ?
4. Quels sont l'opinion et les souhaits des élèves en matière d'évaluation de l'expression orale ?

Les réponses à la question 1 sont basées sur l'observation de leçons et sur les interviews rétrospectives des enseignant-e-s; les réponses à la question 2 s'appuient aussi sur les interviews des enseignant-e-s. La question 3 a été étudiée à la lumière d'épreuves d'expression orale authentiques filmées, tandis que la question 4 a été analysée à l'aide d'un questionnaire adressé aux élèves et d'interviews rétrospectives réalisées avec les élèves qui avaient été filmé-e-s.

3 Phases du projet et méthodologie de recherche

Dans la première phase du projet, l'accent a été mis sur l'évaluation formative. Afin d'étudier les pratiques d'évaluation formative mises en place par les enseignant-e-s et leur attitude face à l'évaluation, deux leçons de DaF ont été observées dans douze classes du secondaire I, suivies d'interviews des enseignant-e-s.¹ Huit enseignant-e-s avec des classes de niveau d'exigence inférieur et moyen de la 8^e et la 9^e année du canton de Fribourg (10H et 11H selon HarmoS) ainsi que quatre classes du canton de Neuchâtel ont participé à l'étude.² Tou-te-s les enseignant-e-s disposaient d'une formation à l'enseignement de l'allemand au niveau secondaire I, avec une expérience dans l'enseignement allant de 2 et 25 ans, soit 7,7 ans en moyenne. La participation à l'étude était volontaire. On peut donc supposer que les enseignant-e-s participant-e-s s'intéressaient au sujet de l'«évaluation (formative) de l'expression orale» et se sentaient relativement à l'aise dans ce domaine. Comme il s'agissait d'observer des cours aussi authentiques que possible, les enseignant-e-s ont simplement été invité-e-s à donner des leçons «axées sur l'expression orale». Les interviews, partiellement standardisées, ont été réalisées à la fin de la deuxième leçon. D'une durée d'environ une heure à chaque fois et articulées en deux parties, elles ont permis de recueillir l'opinion des enseignant-e-s sur l'évaluation de l'expression orale. Dans la première partie, les enseignant-e-s ont été interrogé-e-s sur environ trois situations d'enseignement dans lesquelles des pratiques d'évaluation avaient été observées. Une fois la situation esquissée par l'intervieweuse, les enseignant-e-s ont généralement expliqué spontanément les raisons de leur action, et dans quelles situations celle-ci s'appliquait. La deuxième partie de l'interview a permis de recueillir d'autres pratiques et points de vue des enseignant-e-s sur l'évaluation (formative et sommative) de l'expression orale. Les interviews ont été reproduites selon les règles de la transcription minimale GAT 2 (Selting et al., 2009) puis soumises à une analyse de contenu (Mayring, 2010).

Dans la deuxième phase du projet, l'accent a été mis sur les épreuves sommatives d'expression orale. Dans huit classes secondaires,³ des épreuves authentiques d'expression orale

1. Nous adressons nos remerciements les plus vifs à chaque enseignant-e et chaque élève ayant participé à l'étude pour sa confiance et son ouverture d'esprit. Un grand merci à Thomas Studer également, ainsi qu'à notre groupe d'accompagnement scientifique, pour leurs nombreux retours constructifs sur les différentes étapes du projet.
2. Dans le canton de Neuchâtel, on ne distingue pas encore de niveaux d'exigence dans l'enseignement de l'allemand en 7^e année (9H selon HarmoS). En 8^e et 9^e année (10H et 11H), l'allemand est ensuite enseigné à deux niveaux. Une classe de 7^e année, deux classes à exigences de base et une classe à exigences étendues ont été étudiées à Neuchâtel.
3. Dans le canton de Fribourg, quatre classes ont participé : trois classes du niveau d'exigences le plus bas (deux classes de 8^e (10H) et une classe de 9^e (11H)) et une classe de 8^e (10H) du niveau d'exigences moyen. Dans le canton de Neuchâtel, deux classes ont participé : une 7^e classe (9H, sans attribution de niveau) et une 9^e classe (11H) de niveau inférieur. Dans le canton du Tessin, deux classes ont participé : une 8^e classe (10H) de niveau d'exigence supérieur et une 9^e classe (11H) de niveau d'exigence inférieur.

ont été filmées et suivies d'interviews rétrospectives des enseignant-e-s et des élèves. En outre, les élèves ont rempli un questionnaire sur les épreuves d'expression orale en allemand langue étrangère (pour plus de détails sur les outils d'enquête, voir figure 1). Toutes les épreuves ont été organisées par les enseignant-e-s et se sont déroulées la plupart du temps, à notre demande, dans une salle séparée. La caméra a été placée de manière à ce que les élèves et l'enseignant-e soient visibles sur la vidéo. Pendant les épreuves, l'équipe de recherche ne se trouvait pas dans la salle d'examen, mais s'occupait du questionnaire écrit et des interviews rétrospectives des élèves. Les interviews rétrospectives des enseignant-e-s ont été effectuées, en ligne via MS Teams, plus tard le jour même ou le lendemain.⁴

Suite à la récolte des données, les épreuves d'expression orale et les interviews ont été reproduites à l'aide de la transcription minimale GAT 2 qui inclut les gestes et les expressions faciales. L'analyse conversationnelle qui a suivi s'est concentrée sur le comportement d'intervention des enseignant-e-s durant les épreuves orales en binôme en raison de la variabilité manifeste sur ce point (pour plus de détails, voir Peyer et al., soumis pour publication). Les questions de recherche suivantes se sont révélées essentielles :

1. Lors des épreuves orales par deux, quand et comment les interventions des enseignant-e-s du niveau secondaire I se déroulent-elles ?
2. Quel est l'effet des différents types d'intervention sur l'interaction entre pairs ?

Pour répondre à la seconde question, nous nous basons sur le concept de progressivité (*progressivity*, Schegloff, 2007), qui se réfère à la progression de la conversation et que nous opérationnalisons de la manière suivante :

- a. Les interventions des enseignant-e-s sont propices à la progression de l'interaction entre pairs si celle-ci peut se poursuivre sans aucune autre intervention verbale de l'enseignant-e.
- b. Les interventions sont d'autant moins propices à la progression de l'interaction entre pairs que d'autres interventions verbales de l'enseignant-e sont nécessaires avant que l'interaction entre pairs puisse se poursuivre, ou lorsque les candidat-e-s doivent poser des questions complémentaires.

Les interviews des enseignant-e-s et des élèves ont fait l'objet d'une analyse de contenu (Mayring, 2010). Le questionnaire destiné aux élèves comportait des questions fermées et ouvertes. Les questions ouvertes ont également fait l'objet d'une analyse de contenu (pour plus de détails sur l'enquête auprès des élèves, voir Lüthi et al., soumis).

	Méthode	Participant-e-s
Phase 1 du projet mars-juin 2021	<i>Observation de leçons</i> Consigner l'évaluation formative, notamment les commentaires des enseignant-e-s	12 classes/enseignant-e-s à 2 leçons (8 de Fribourg, 4 de Neuchâtel)
	<i>Interviews rétrospectives des enseignant-e-s</i> Partie 1: (rétrospective) sur des situations d'évaluation observées précédemment, partie 2: questions générales sur l'évaluation	12 enseignant-e-s
Phase 2 du projet mars-juin 2022	<i>Vidéographie d'épreuves d'expression orale authentiques</i> Épreuves créées et organisées par l'enseignant-e (épreuves orales par deux et individuelles)	82 élèves: 48 épreuves dans 8 classes (4 de Fribourg, 2 de Neuchâtel, 2 du Tessin)
	<i>Interviews rétrospectives des enseignant-e-s</i> (Thèmes: évaluation de l'épreuve/ des compétences de la classe, préparation, évaluation/notation, utilité des épreuves orales)	7 enseignant-e-s (4 de Fribourg, 2 de Neuchâtel, 1 du Tessin avec 2 classes)
	<i>Interviews rétrospectives des élèves</i> Brèves interviews d'élèves filmé-e-s (thèmes: autoévaluation, préparation, utilité)	Interviews de 84 élèves
	<i>Questionnaires élèves</i> (Thèmes: préparation de l'épreuve, peur de l'examen, feedback, utilité des différents formats d'épreuve, autoévaluation)	254 élèves (17 classes) (Fribourg: 57%, Neuchâtel: 13%, Tessin: 30%)

Figure 1: Aperçu des récoltes de données

4 Résultats

Les principaux résultats des deux phases du projet sont résumés ci-dessous.

4.1 Quelles sont les pratiques d'évaluation formative utilisées par les enseignant-e-s du secondaire I et comment sont-elles explicitées ?

Dans le cadre des 24 leçons observées, les enseignant-e-s ont principalement proposé des activités d'expression orale structurées (travail avec des plans de dialogues, réponse à des questions, jeux), dont le but premier était de faire pratiquer de nouveaux mots et structures grammaticales. On a pu constater que les enseignant-e-s donnaient essentiellement un feedback spontané et correctif. Des pratiques d'évaluation formative ou de feedback planifiées et plus formelles n'ont été que rarement observées. La forme de feedback correctif la plus fréquemment observée a été le *recast*, c'est-à-dire la reformulation correcte par l'enseignant-e de propos erronés (Lyster & Ranta, 1997). Lors des interviews, certain-e-s enseignant-e-s ont dit apprécier les *recasts* précisément parce que c'est un moyen de ne pas dire explicitement qu'il y a une erreur, et – selon l'hypothèse de deux enseignant-e-s – de permettre aux apprenant-e-s de se rendre compte de leurs propres erreurs. Les *prompts* (différentes indications qui incitent à l'autocorrection, Lyster & Saito, 2010) ont également été utilisés régulièrement, mais pas fréquemment, par les enseignant-e-s. Dans quelques classes, on a pu observer que des pairs apportaient des corrections et des explications pendant une activité d'expression orale. Une enseignante a déclaré en interview encourager la correction par les pairs parce que les élèves ont une façon différente de donner des explications et qu'elle a l'impression qu'elles sont mieux comprises – une appréciation partagée par un-e collègue. On a également relevé que quatre enseignant-e-s avaient incité la classe ou un-e élève à s'autoévaluer (spontanément). Par exemple, après qu'une activité d'expression orale ait été préparée individuellement, un-e enseignant-e demanda à un élève s'il était en mesure de parler cinq minutes du sujet de l'activité. La réponse de l'élève a été « clairement pas ». En posant cette question, l'enseignant-e voulait – selon ses déclarations dans l'interview – que les apprenant-e-s se rendent compte qu'ils et elles devaient encore mieux apprendre le vocabulaire.

Les interviews ont montré que les enseignant-e-s éprouvent généralement des difficultés à donner un feedback sur les compétences orales, par manque de temps et en raison de la taille des classes. De plus, les enseignant-e-s craignent de démotiver les apprenant-e-s en les corrigeant. Les corrections ou les feedbacks sont donc souvent adressés à l'ensemble de la classe et ne sont pas individualisés. De même, tou-te-s les enseignant-e-s indiquent lors des interviews ne pas corriger délibérément dans certaines situations. La raison la plus souvent invoquée est un niveau de compréhension suffisant. Ce que l'on entend exactement par-là dépend de la situation. En cas d'erreur avérée, un certain nombre

d'enseignant-e-s déclarent se demander si le propos erroné resterait compréhensible pour un-e locuteur-trice de langue maternelle allemande et corrigent – ou non – en fonction de cela. La seconde raison la plus souvent citée est la motivation. C'est ce que dit l'enseignant-e 07 : « J'ai pas envie de les corriger pour tout en fait [...] je peux pas toujours les casser ». La motivation a été mentionnée de manière récurrente lors des interviews. Selon les enseignant-e-s, il n'est pas facile de faire parler les élèves. Ils et elles supposent que cela est dû à un manque de vocabulaire, la peur de faire des erreurs ou de s'exposer. C'est pourquoi, par crainte d'un effet démotivant, on veille à ne pas (trop) corriger. La plupart du temps, les enseignant-e-s insistent sur le fait que les apprenant-e-s doivent être encouragés à parler allemand. À cet égard, les erreurs sont secondaires : « c'est plus important pour moi qu'il parle allemand que qu'il parle JUSTE » (enseignant-e 03). Les éloges et les compliments sont également utilisés délibérément dans le but de renforcer la motivation des apprenant-e-s ou leur estime de soi. En classe, on a surtout observé de brefs commentaires tels que « très bien » (pour d'autres résultats de la première phase du projet, voir Peyer et al., 2023).

L'objectif général de maintenir la motivation permet probablement d'expliquer la tendance des enseignant-e-s à privilégier les types de correction plus implicites que sont les *recasts*. À l'instar des enseignant-e-s dans d'autres études (Kamiya, 2016; Roothoof, 2014), une partie des enseignant-e-s interviewés ont indiqué craindre que leurs corrections ne provoquent des réactions affectives négatives chez les apprenant-e-s. La méta-analyse de Lyster & Saito (2010) sur les études en salle de classe a cependant montré que les *recasts* se révèlent moins efficaces que les *prompts*, et qu'il serait donc souhaitable d'utiliser davantage ces derniers. Cela étant dit, il est important de rester attentif à la motivation des élèves et à leurs réactions affectives face à un feedback correctif. Comme le note Roothoof (2014), il est étonnant que si peu d'études se soient penchées sur les réactions affectives des apprenant-e-s aux corrections, alors que les enseignant-e-s ont manifestement leurs propres hypothèses à ce sujet, lesquelles influencent à leur tour quel type de corrections sont utilisés.

4.2 Comment les épreuves orales par deux se déroulent-elles au niveau secondaire I ?

Comme mentionné ci-dessus, des tests d'expression orale authentiques ont été filmés dans huit classes. Les épreuves orales se déroulaient soit par deux (4 classes), soit en individuel (1 classe), soit en deux parties, l'une en binôme et l'autre en individuel (3 classes). Nous allons examiner ici plus en détail le comportement adopté durant l'examen par quatre enseignant-e-s ayant fait passer les épreuves orales en binôme ($n=21$). Ce paramètre est nettement plus important pour des élèves de niveau A1-A2, ce qui était le cas de la plupart des élèves filmés, que pour des apprenant-e-s plus avancés. Selon le Cadre européen commun de référence pour les langues (CECR), en effet, les apprenant-e-s au niveau A2 sont encore à peine capables d'entretenir une conversation (Conseil de l'Europe, 2001: 79).

Selon Galaczi, les épreuves orales en binôme au niveau A2 nécessitent donc encore un soutien et un contrôle réguliers de la part des gestionnaires de l'examen (Galaczi, 2014 : 557).

Activités d'examen et déroulement des épreuves

Les activités d'examen ont été élaborées par les enseignant-e-s concerné-e-s, le plus souvent en collaboration avec des collègues de leurs écoles. La figure 2 donne un aperçu des enseignant-e-s, des niveaux de classe et des types d'examens filmés.

Enseignant-e	Eva ⁵	Mia	Lea	Tom
Expérience dans l'enseignement	4 ans	25 ans	7 ans	20 ans
Classe	Classe à exigences de base 8 ^e année (10H)	Classe à exigences de base 8 ^e année (10H)	Classe générale 8 ^e année (10H)	Classe à exigences de base 9 ^e année (11H)
Activité d'évaluation	Poser 5 questions sur le thème de la mode et y répondre		Poser 5 questions sur le thème de l'argent de poche et y répondre à l'aide d'une fiche signalétique	Organiser une activité en commun (fête, excursion)
Objectif de l'épreuve	Vérification de l'acquisition des objectifs d'apprentissage du chapitre du manuel			Préparation à l'examen final cantonal

Figure 2 : Informations sur les enseignant-e-s, les classes et les types d'examen

Eva, Mia et Lea ont effectué des tests pour vérifier l'acquisition des objectifs d'apprentissage d'un chapitre du manuel (*achievement test*), et Tom un test de préparation à l'examen final cantonal (*proficiency test*). Les activités d'évaluation (y compris les instructions) des quatre enseignant-e-s étaient rédigées en allemand. Eva et Mia ont préparé leur activité d'évaluation ensemble, car elles enseignaient dans la même école. Dans leur test, les élèves

5 Les noms propres (de personnes et de lieux mentionnés dans les transcriptions) ont été pseudonymisés.

devaient parler de mode, poser chacun-e au moins cinq questions à ce sujet et y répondre. À titre d'aide, plusieurs pronoms interrogatifs et quelques verbes étaient imprimés en allemand sur la feuille d'examen. Dans l'épreuve de Lea, les élèves devaient parler d'argent de poche, poser cinq questions sur le sujet et y répondre. Pour ce faire, ils et elles ont reçu une fiche signalétique d'un-e jeune, qui contenait notamment des informations (le plus souvent sous forme d'images) sur le montant d'argent de poche reçu, le job d'étudiant-e et l'objectif d'épargne. De même, quelques expressions comme «verdienen» [gagner], «arbeiten» [travailler] ou «genug/zu wenig» [assez/trop peu] étaient imprimées en allemand. Dans l'épreuve de Tom, les élèves devaient organiser une activité à deux (sortie au cinéma, à la piscine, etc.) et discuter de différents points tels que l'heure, le moyen de transport, etc. Sa feuille d'examen contenait également quelques expressions en allemand. Les enseignant-e-s ont toutes et tous accordé aux élèves un temps de préparation, sans toutefois les autoriser à prendre des notes. Les épreuves filmées duraient entre trois et six minutes.

Lors des épreuves orales (par deux), quand et comment les interventions des enseignant-e-s du niveau secondaire I se déroulent-elles ?

Les quatre enseignant-e-s observé-e-s se comportent de manière très différente lors des épreuves orales par deux. Ainsi, on constate une grande variabilité entre les enseignant-e-s, tant au niveau de la fréquence de leurs interventions (de rares pour Tom à fréquentes pour Lea) que de la diversité de leurs fonctions : Tom n'intervient pas dans une épreuve, même en cas de problèmes d'interaction évidents, sauf pour rappeler aux élèves de résumer leur discussion à la fin d'un examen. En revanche, les trois autres enseignantes interviennent non seulement dans le but de gérer l'examen (par ex. inviter un-e élève à poser une question), mais aussi pour aider les candidat-e-s à résoudre des problèmes d'interaction (difficultés de compréhension, recherche de mots, etc.) et pour leur donner un feedback positif (généralement non verbal). Mia et Lea interviennent aussi parfois durant une épreuve pour corriger les propos d'un-e candidat-e, généralement par une question de clarification ou un *recast*. De plus, Lea intervient plusieurs fois durant une épreuve pour vérifier si l'élève est capable d'utiliser des structures que la classe a apprises récemment, comme les propositions subordonnées causales ou finales. L'une des raisons pour lesquelles Lea essaie toujours d'éliciter des structures grammaticales pendant les épreuves est sans doute l'interdiction faite au corps enseignant fribourgeois depuis quelques années de faire passer des tests de grammaire et de vocabulaire dans les cours de langues étrangères. En effet, les connaissances en grammaire et en vocabulaire doivent être vérifiées en même temps que les compétences linguistiques communicatives (SEnOF, 2018).

Dans l'ensemble, la variabilité considérable du comportement des enseignant-e-s est probablement due à divers facteurs, comme des différences d'objectif ou d'activités d'examen utilisées. Ce n'est probablement pas un hasard si seul Tom, qui intervient peu dans les épreuves, a recours à une activité d'évaluation qui ressemble à une tâche, du moins à certains égards (par ex. en ce qui concerne l'orientation vers un objectif ou l'authenti-

cité), et qui rappelle fortement les tâches d'examen que l'on rencontre dans les tests de *proficiency* standardisés pour le niveau A2. Dans les activités des autres enseignantes, par exemple, le *turn-taking* (tour de parole) est réglé à l'avance, elles ne font pas appel à un usage naturel de la langue et n'encouragent pas la négociation collective – des aspects qui, dans la littérature, sont considérés comme des avantages des épreuves orales en binôme (par ex. May, 2009; Rydell, 2019). L'activité de Lea, en particulier, présente un certain manque de clarté (notamment une photo qui prête à confusion sur la feuille d'examen), ce qui l'oblige à répondre régulièrement à des questions sur l'activité pendant l'épreuve. En outre, la personnalité de l'enseignant-e semble jouer un rôle important dans la fréquence et la manière dont il ou elle intervient dans les épreuves orales par deux, comme en témoigne la différence de comportement entre Eva et Mia, qui ont utilisé la même activité avec des élèves de même niveau.

Quel est l'effet des différents types d'intervention sur l'interaction entre pairs ?

Dans cette section, nous analysons les stratégies d'intervention typiques et récurrentes en fonction de leur impact sur la progression de l'interaction entre pairs, en commençant par celles qui la favorisent et en terminant par celles qui la ralentissent.

Selon nos observations, l'une des interventions les plus efficaces pour soutenir l'interaction entre pairs est la reformulation d'une question difficile à comprendre, une sorte de stratégie de médiation que Mia utilise régulièrement avec succès. L'extrait 1 en est une parfaite illustration.

- 01 E1 : äm:: (6.1) wo (2.9) wo: (2.8) <non> (1.3) ((chuchote non-int., env. 1 sec))
euhm où où :
- 02 weIche (1.3) trage ((tape sur la table avec l'index à chaque mot)) (-)
quel porte
- 03 schwarz (-) t-shirt ((regarde E2))
noir t-shirt
- 04 (1.0)
- 05 Mia : ((tourne son regard vers E2)) trägst du schwarze t-shirts
portes-tu des t-shirts noirs
- 06 E2 : äh ja ich: äh: trage (-) äh schwarze t-shirt
euh oui je: euh: porte euh t-shirt noir

Extrait 1: Reformulation d'une question difficile à comprendre (Mia, épreuve 5)

L'élève 1 (E1) essaie de formuler une question avec le pronom interrogatif «où» à la ligne 1, mais abandonne et commence à en former une autre avec «quel» à la ligne 2. Le sens de sa question n'est pas clair (quel porte t-shirt noir?), si bien que Mia intervient et pose directement à l'élève 2 (E2) la question suivante: tu portes des t-shirts noirs? L'élève 2 peut alors répondre à cette question sans plus d'hésitation. Reste à savoir si la reformulation de Mia correspond à la question que l'élève 1 souhaite réellement poser et si son intervention

était vraiment nécessaire, ou si les candidates auraient été en mesure de résoudre elles-mêmes leurs (possibles) difficultés d'interaction, avec, par exemple, l'élève 2 demandant des éclaircissements à l'élève 1.

Une autre stratégie d'intervention qui s'est toujours avérée favorable à la progression de l'entretien lors des examens observés est l'invitation de l'enseignant-e à improviser ou à s'écarter de la fiche signalétique. Lea utilise cette stratégie de temps en temps, lorsqu'un-e candidat-e cherche ses mots ou fait comprendre qu'elle ou il a des difficultés de formulation. Dans l'extrait 2, l'élève 7 (E7) a du mal à formuler que son travail d'appoint est de faire des ménages, comme on peut le voir sur sa fiche signalétique, si bien que Lea l'encourage à « simplement donner une information » (ligne 4).

- 01 E8 : hm und in was jobbst du ((regarde E7))
hum et dans quoi tu travailles
- 02 E7 : ich jobbe in äh (1.9) äh: <ah (non) ça je sais pas dire> (2.7) ah (1.0)
je travaille dans euh euh:
- 03 Lea : du musst nicht GENAU sagen was auf dem foto ist du kannst a:uch
tu n'es pas obligé de dire EXACTEMENT ce qu'il y a sur la photo tu peux au:ssi
- 04 ((mouvement circulaire de la main)) (---) einfach eine information geben
simplement donner une information
- 05 E7 : okay ich jobben in (1.5) äh ich <help>
ok je travailler en euh je <aide>
- 06 ((regarde Lea, se gratte la tête)) [mein jobben ist zu helfen]
[mon job est d'aider]
- 07 Lea : [((regarde E7, hoche la tête plusieurs fois))]
- 08 E7 : (-) äh neuest (-) äh: (---) neuest schu:\ (-) sch\ sch: öh schule
euh plus nouvelle euh: plus nouvelle éc:\ è\ é: euh école
- 09 Lea : ((hoche la tête) hm hm
- 10 E7 : öh in der <co> <of> fribourg (-) mit (-) den mathematik und französisch
euh en <co> <of> fribourg avec le mathématique et français
- 11 Lea : sehr gut (-) fantastisch
très bien fantastique

Extrait 2: Invitation à improviser (Lea, épreuve 4)

Comme le montre l'extrait 2, l'élève 7 (E7), après avoir été invité à improviser, peut dire en termes simples, de manière hésitante et avec des recours occasionnels à l'anglais et au français («help» au lieu de *helfen* [aider] à la ligne 5 et à la ligne 6; «of» et «CO» pour cycle d'orientation, c'est-à-dire école secondaire, à la ligne 10), mais sans autre aide, qu'il donne des cours d'appui en mathématiques et en français dans une école secondaire. Il suffit pour cela que l'enseignante hoche la tête en signe d'encouragement pendant qu'il cherche ses mots. Une invitation à improviser permet donc de vérifier ce que les candidat-e-s peuvent exprimer avec des moyens linguistiques très limités et éventuellement de les aider à réussir. Par ailleurs, l'extrait 2 montre que les attentes concernant la réalisation de l'activité ne sont

pas claires : en permettant à cet élève de produire librement au lieu de reproduire des contenus donnés, Lea simplifie l'activité de test. Cette aide devrait alors être prise en compte dans l'évaluation.

Les interventions qui sont parfois propices à la progression de l'interaction entre pairs sont les demandes de clarification dans la langue de scolarisation, une stratégie que Mia utilise régulièrement.

- 01 E3 : ö:m (5.8) ä:m (1.2) wo ist (1.1) äh die shoppen
euhm euh:m où est euh le faire du shopping
- 02 (0.9)
- 03 Mia : ((regarde E3))
- 04 E3 : ((contact visuel avec Mia)) äm (1.5)
euhm
- 05 Mia : <hm qu'est-ce que t'aimerais dire [S3]>
- 06 E3 : [äh]
[euh]
- 07 E3 : <où il va faire shop\ le shopping>
- 08 Mia : oh wo shoppst du ((regarde E4)) [wo gehst du] shoppen
oh OÜ fais-tu du shopping [où vas-tu] faire du shopping
- 09 E3 : [ah wo shopp\]
[ah où shopp\]
- 10 E4 : ä:m (2.3) ich shoppe ((regarde Mia qui regarde la grille, 1.0 sec.))
euhm: je fais du shopping
- 11 ((regarde dans le vide, 0.8 sec.) äh: ((respire profondément)) (-) in fribourg
euh: à fribourg

Extrait 3: Question de clarification dans la langue de scolarisation (Mia, épreuve 4)

Dans l'extrait 3, l'élève 3 (E3) formule, après une longue hésitation, une question erronée et difficile à comprendre avec le pronom interrogatif «où» (ligne 1). Mia le regarde alors (ligne 3) et, après une courte pause, passe au français pour demander ce qu'il voulait dire, une question à laquelle l'élève 3 répond également en français. Mia formule alors une question correcte en allemand (ligne 8), que l'élève 3 commence immédiatement à répéter à la ligne 9. L'intervention de Mia conduit donc à un *learner uptake* partiel (Lyster & Ranta, 1997 : 49). Son camarade, l'élève 4 (E4), donne finalement une réponse hésitante correcte, sans avoir à demander lui-même de l'aide. Mia utilise régulièrement des questions de clarification, bien qu'elles ne permettent de résoudre (rapidement) un problème d'interaction que dans la moitié des cas environ. De même, comme le montre l'extrait 3, il ne suffit pas de demander à l'élève ce qu'il ou elle veut dire en utilisant simplement la langue de scolarisation. Dans un deuxième temps, l'enseignant-e doit traduire la réponse en allemand. Ces questions de clarification indiquent en outre que les élèves ne disposent pas encore des stratégies nécessaires pour résoudre eux- ou elles-mêmes leurs difficultés d'interaction, ce que l'on attend généralement des apprenant-e-s de niveau A2 (Council of Europe, 2018 : 84).

De manière générale, les élèves observé-e-s semblent avoir peu de stratégies à leur disposition, notamment peu de stratégies de compensation lors de la recherche de mots.

Selon nos observations, certaines interventions, notamment l'éllicitation de mots et de structures, ne favorisent pas la progression de l'interaction entre pairs. Dans l'extrait 4, l'élève 7 (E7) doit dire comment il dépense son argent de poche en utilisant les informations de la fiche signalétique, sur laquelle l'image correspondante montre un cadeau, des pommes frites et une saucisse.

- 01 E7 : (--) öm (-) ich kaufen mein gelb für geschenke ün öh geschenke für mein freundIN
euhm j'acheter mon argent pour cadeaux et euh des cadeaux pour ma amie
- 02 Lea : ((sourit et hoche la tête))
- 03 E7 : (-) und öh (1.7) äh: (1.2) <saucisse> ((lève son regard vers Lea))
et euh euh:
- 04 Lea : ((regarde E7, 1.0 sec.)) was ist das wort auf deutsch für <saucisse>
quel est le mot en allemand pour <saucisse>
- 05 ((geste avec les paumes vers le haut))
- 06 E7 : ah <mais je [sais]>
- 07 Lea : [oder kennst] du [öh]
[ou connais] tu [euh]
- 08 E7 : [[<c'est>]] euh frü\ frü\ früst ((regarde Lea))
- 09 <non> (--) ((regarde la feuille))
- 10 Lea : po
- 11 E7 : ((regarde Lea))
- 12 Lea : (---) pom
- 13 E7 : (-) pom f[rüst]
- 14 Lea : [pommes]
[frites]
- 15 E7 : pommes
frites
- 16 Lea : pommes ((hoche la tête)) okay ((regarde E8))
frites ok

Extrait 4: Élicitation d'un mot (Lea, épreuve 4)

Comme l'élève 7 (E7) ne se souvient pas du mot allemand pour « saucisse » (qui serait «Wurst»), il utilise le mot français et regarde l'enseignante en cherchant de l'aide (ligne 3). Lea demande alors quel est le mot allemand pour « saucisse » (ligne 4). Mais comme l'élève 7 ne peut justement pas le nommer (tout de suite), elle essaie ensuite d'élucider les frites également représentées sur l'assiette en demandant « ou tu connais » (ligne 7). Cependant, l'élève 7 semble toujours chercher le mot pour « saucisse », car il répond à cela : « c'est euh frü\ frü\ früst » (ligne 9), ce qui rappelle le pluriel « Würste ». Lea ne semble pas l'avoir en-

tendu, car elle continue d'éliciter le mot « pommes frites » en disant la première syllabe du mot. En fin de compte, il faut trois interventions de Lea pour que l'élève 7 soit finalement capable de dire ou de répéter le mot « pommes frites ». Cet exemple montre que l'élicitation d'un mot par l'enseignant-e dans le cas d'une recherche de mots n'est pas forcément utile à la progression de l'interaction. Au contraire, l'intervention de Lea ci-dessus a interrompu l'interaction entre les deux candidat-e-s pendant un certain temps et a sans doute aussi perturbé l'élève 7.

Plus souvent que des mots isolés, Lea élicite des structures complexes, par exemple des propositions subordonnées avec *damit* [pour que]. Cela interromp toujours l'interaction entre pairs pendant une longue période, car il faut au moins deux interventions et, la plupart du temps, une correction par l'enseignant-e. Le but premier de Lea est sans doute de vérifier si les candidat-e-s sont capables d'utiliser une structure qui vient d'être étudiée en classe. Cette intervention ne vise donc pas à garantir la poursuite de l'interaction entre les deux candidat-e-s et ne sert pas l'un des objectifs principaux des épreuves orales par deux, à savoir la vérification des compétences d'interaction.

4.3 Quelle est l'utilité que les enseignant-e-s voient dans l'évaluation sommative de l'expression orale ?

Au cours de la deuxième phase de recherche, il a été demandé au sept enseignant-e-s, lors des interviews rétrospectives, comment seraient utilisées les connaissances acquises dans le cadre de l'épreuve (« Qu'allez-vous faire avec les informations que vous avez obtenues grâce à cette épreuve ? »). La plupart des enseignant-e-s ont commencé par aborder la question du feedback donné aux élèves après l'épreuve. Une partie des enseignant-e-s rendent aux élèves la grille remplie avec la note (et éventuellement quelques commentaires sur les erreurs), mais tous et toutes doutent que les élèves utilisent ce feedback dans leurs apprentissages. C'est pourquoi une enseignante explique oralement aux élèves comment le nombre de points a été calculé et la note a été obtenue. Deux autres enseignant-e-s, qui ont également des doutes que les élèves intègrent un feedback écrit, donnent à chaque élève la possibilité d'écouter ensemble l'enregistrement audio de l'épreuve afin d'obtenir un feedback oral individuel. D'autres enseignant-e-s encore ne donnent pas non plus de feedback écrit aux élèves, mais se tiennent à leur disposition pour répondre à leurs questions (sur la note ou la performance individuelle). En plus du feedback individuel, une partie des enseignant-e-s déclarent transmettre à toute la classe un feedback collectif sur les erreurs fréquemment observées. Une partie des enseignant-e-s indique aussi encourager la classe ou les élèves plus anxieux-euses lors du feedback, voire de tenter de renforcer leur confiance en soi. Deux enseignant-e-s précisent en outre utiliser les informations obtenues pour réviser l'épreuve elle-même ou la grille d'évaluation. Un autre a décidé spontanément, en cours d'examen, de ne pas noter l'épreuve mais de la refaire et d'encore mieux y préparer les élèves.

Il a aussi été demandé aux sept enseignant-e-s si, à leur avis, les épreuves d'expression orale étaient utiles aux élèves pour progresser en allemand. Les enseignant-e-s ne s'ac-

cordent pas sur cette question, les réponses étant pour moitié négatives et pour moitié positives. Ainsi, Lea ne voit généralement pas l'utilité des épreuves orales : « c'est plus une source de stress ces évaluations que une source de progression ». Pour motiver les élèves à apprendre, il serait beaucoup plus efficace d'organiser une rencontre avec une classe partenaire de Suisse alémanique. Une autre enseignante exprime également des doutes quant à l'utilité des examens oraux pour ses élèves, car à leur âge, ils et elles ne voient pas encore l'intérêt d'apprendre l'allemand. Deux autres enseignant-e-s font une distinction entre le niveau d'exigence inférieur et le niveau d'exigence supérieur. A leur avis, les épreuves orales n'apportent rien aux classes les plus faibles ayant participé à l'étude, car ces élèves ne se préparent de toute façon pas en vue d'un examen alors que les classes ayant un niveau d'exigence plus élevé le font. Les trois autres enseignant-e-s ont une attitude plus positive vis-à-vis des épreuves orales. Eva et Tom, qui ont tous deux participé à l'étude avec une classe à niveau d'exigence inférieur, considèrent les épreuves avant tout comme un moyen (de pression) pour inciter les élèves à travailler avec plus de motivation ou à s'exercer avec plus de sérieux. Un autre enseignant raisonne différemment et pense que, de tous les examens, les épreuves orales sont celles qui font faire le plus de progrès : en effet, les élèves prennent conscience de leur niveau et de leur capacité à interagir dans une situation quotidienne. Cela les renforce et leur permet de se rendre compte des efforts qu'il leur faut encore fournir. Cet enseignant apprécie aussi les épreuves orales parce qu'il peut aider les élèves plus directement que lors des épreuves écrites et parce qu'elles sont l'occasion de donner un feedback verbal aux élèves immédiatement après l'épreuve.

Dans l'ensemble, les enseignant-e-s interviewé-e-s se montrent quelque peu résigné-e-s quand on les interroge sur l'utilité des épreuves orales, particulièrement la pérennité de leur feedback, et quelques peu perplexes sur comment ils et elles pourraient améliorer leur feedback. La plupart ne pensent pas que le feedback écrit (par ex. la grille remplie) soit assimilé par les élèves ou utilisé pour la suite de l'apprentissage. Il ressort de la recherche que le feedback est particulièrement utile pour l'apprentissage lorsqu'il est donné rapidement, qu'il est concret et descriptif, et qu'il montre (ce) que les apprenant-e-s peuvent (s')améliorer et comment (entre autres Luoma, 2004; Hattie & Timperley, 2007). Ce type de feedback concret (oral), suivant immédiatement l'épreuve, n'a pu être observé que chez deux des enseignant-e-s, probablement pour des raisons pratiques et temporelles. Il n'est donc pas étonnant qu'environ la moitié des sept enseignant-e-s interviewé-e-s soient, pour diverses raisons, plutôt sceptiques quant à l'utilité des épreuves d'expression orale notamment dans les classes de niveau d'exigence inférieur.

4.4 Quels sont les besoins des enseignant-e-s en matière d'évaluation sommative de l'expression orale ?

Les interviews ont montré que les enseignant-e-s se heurtent à plusieurs difficultés lorsqu'il s'agit de faire passer des épreuves d'expression orale. Ainsi, elles prennent beaucoup de temps, surtout lorsque les enseignant-e-s enregistrent les prestations orales et ne procèdent à l'évaluation qu'ensuite. En outre, des questions de discipline ont été relevés à plusieurs reprises, qui empêchent les enseignant-e-s d'organiser les examens en dehors de la salle de classe, les élèves ne pouvant être laissés-e-s sans surveillance. Nous avons donc demandé aux douze enseignant-e-s de la première phase si ils et elles souhaitaient recevoir de l'aide ou du matériel supplémentaire pour l'évaluation sommative de l'expression orale. À cette question, les réponses des enseignant-e-s de Fribourg et de Neuchâtel diffèrent sensiblement, ce qui s'explique probablement par des contextes différents. Dans le canton de Fribourg, l'organisation de plusieurs épreuves orales par an (en fonction du nombre d'heures hebdomadaires) est obligatoire, de même que l'utilisation d'une grille d'évaluation spécifique qui distingue quatre aspects (Tâche, Communication, Aspects formels, Prononciation et intonation). De son côté, le canton de Neuchâtel recommande simplement à ses enseignant-e-s d'organiser des épreuves orales, mais ne donne aucune indication quant à la fréquence ou la nature de la grille d'évaluation à utiliser. Lorsqu'on leur demande ce qui pourrait les aider, les quatre enseignant-e-s de Neuchâtel soulignent à plusieurs reprises l'utilité du matériel d'évaluation que la Haute école pédagogique met à leur disposition. Tout autre matériel spécifique et directement exploitable pour l'évaluation de l'expression orale ou du matériel pouvant être adapté à différents groupes resterait néanmoins bienvenu.

Interrogé-e-s sur les aides à l'évaluation, les huit enseignant-e-s de Fribourg expriment spontanément leur opinion sur la grille obligatoire. Quatre d'entre eux-elles se sont montrés-e-s très critiques à l'égard de cet outil, deux enseignant-e-s en critiquent certains aspects, tandis que deux autres en sont satisfait-e-s. Les enseignant-e-s critiques s'accordent à dire que la grille n'est pas suffisamment différenciée et qu'elle devrait être affinée. Pour une partie d'entre eux-elles, l'interprétation des catégories est délicate: «erfüllt/nicht erfüllt beim Raster – was heisst das konkret?» (Eva), voire subjective: «Des fois j'ai l'impression que c'est une appréciation très personnelle finalement» (enseignant-e 06), ou «On évalue théoriquement tous de la même façon. Voilà je dis bien théoriquement parce que je suis pas tout à fait convaincue de ça» (enseignant-e 04). Une partie des enseignant-e-s formule aussi des critiques plus fondamentales, comme Lea, qui trouve que la grille ne rend pas justice aux élèves et qu'elle les évalue trop en fonction de leurs lacunes:

J'aime pas vraiment travailler avec ces grilles cantonales. Je trouve que je considère pas assez l'élève et que je me concentre sur ces quatre critères et que chaque fois qu'il y a un truc qui coince je me dis ah ben là j'enlève un point. Puis c'est vraiment j'ai l'impression une évaluation vers le bas plutôt que de valoriser ce qu'ils ont déjà su [...]. J'ai pas l'impression que c'est le meilleur outil qu'on ait pour évaluer l'oralité.

La plupart des enseignant-e-s fribourgeois-es interviewé-e-s souhaitent que la grille soit plus concrète et mieux affinée, qu'elle soit accompagnée d'exemples d'évaluation ou qu'elle ne soit plus obligatoire. Pour le reste, ils et elles aimeraient – si tant est qu'il en existe – davantage de matériel spécifique et différencié pour *exercer* l'expression orale de manière variée.

Nous estimons que les critiques formulées par plusieurs enseignant-e-s à l'encontre de la grille fribourgeoise sont fondées, car en plus de manquer de précision à différents égards, elle présente d'autres défauts. Par exemple, la description des critères laisse une grande marge d'interprétation dans la mesure où la grille ne fait référence ni aux descriptions de compétences du CECR ni au plan d'études et où il n'y a pas d'exemples de référence (c.-à-d. enregistrements d'élèves, commentés et classés). Sans ces références, les gradations pour le critère «Prononciation et intonation» par exemple («En grande partie appropriées, Suffisamment appropriées, Seulement en partie appropriées, À peine appropriées, Pas appropriées») ne peuvent être appliquées qu'arbitrairement. Rien d'étonnant donc à ce que, confronté-e-s à la difficulté d'évaluer les élèves sur ce critère, trois enseignant-e-s déclarent spontanément en interview accorder le même nombre de points (suffisant) à toute la classe. Pour l'expliquer, ces trois enseignant-e-s invoquent l'imprécision de la grille mais aussi le fait que les élèves ne parlent tout simplement pas assez pour que leurs compétences puissent être évaluées. Un autre point faible de la grille est le regroupement, sous le critère «Communication», d'aspects aussi divers que la «Variété de la structure de la phrase», la «Variété du vocabulaire», la «Réalisation de l'intention de parole» y compris l'«Adéquation du vocabulaire au contexte» ainsi que la «Fluidité de la parole» – des aspects qui, dans la grille d'évaluation de l'expression orale de Lingualevel par exemple, sont répartis dans les catégories «Spectre», «Exactitude» et «Fluidité» (Lenz & Studer, 2007). Aux critiques faites à l'encontre de la grille elle-même, s'ajoute le fait que plusieurs enseignant-e-s ont souligné n'avoir reçu que peu ou pas de formation à l'utilisation de la grille, ce qui met encore davantage en question la fiabilité de cet instrument.

4.5 Que pensent les élèves des épreuves d'expression orale ?

Dans le cadre d'un sondage écrit ($n=254$) et d'interviews rétrospectives ($n=84$), les élèves ont également exprimé leur opinion et leurs souhaits concernant les épreuves d'expression orale. Les plus importants sont présentés ici (pour plus d'informations sur le point de vue des élèves, voir Lüthi et al., soumis).

Dans le questionnaire, les élèves se sont exprimés-e-s sur l'utilité des différents formats d'examen pour améliorer leur allemand (figure 3). Il apparaît que globalement les élèves trouvent différents formats utiles. Les épreuves orales en binôme sont considérées comme le format le plus utile (80% d'élèves (plutôt) d'accord), devant les épreuves orales individuelles et les exposés. Dans les interviews, nous avons également demandé aux élèves si les épreuves orales les aidaient à progresser en allemand. À cet égard, 57% ($n=48$) et 23% ($n=19$) des élèves ont jugé ces épreuves utiles ou au moins partiellement utiles, tandis que

18% des élèves ($n=15$) les ont jugées inutiles. Les élèves ont estimé en effet que les épreuves d'expression orale leur permettaient de développer leurs compétences à l'oral (14 élèves, 17%), qu'elles et ils auraient besoin de parler allemand plus tard dans la vie (par ex. apprentissage, travail) ou lors de voyages dans des régions germanophones (10 élèves, 12%), ou encore que cela leur permettait de s'entraîner à la prononciation (9 élèves, 10%). De nombreux arguments convergent donc vers le fait que les épreuves d'expression orale sont une occasion de s'exercer à parler ou de développer certains aspects de la compétence orale, ou qu'il est important de savoir parler et que les épreuves orales sont utiles de ce point de vue. Les élèves pour qui les épreuves d'expression orale sont inutiles ont avancé plusieurs raisons, la plus fréquente étant celle du stress engendré (6 élèves, 7%).

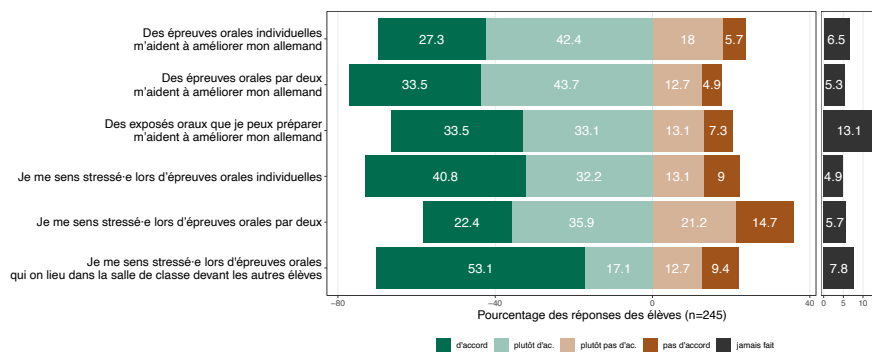


Figure 3: Résultats du questionnaire des élèves sur l'utilité des épreuves et le sentiment de stress en fonction du format d'examen

De même, les élèves devaient indiquer dans le questionnaire si les différents formats d'épreuves orales les stressaient (figure 3). Il est intéressant de noter que les élèves considèrent différents formats d'examen comme stressants, bien qu'utiles, l'épreuve orale en binôme étant perçue à la fois comme le format le moins stressant et le plus utile. Alors que les trois quarts des élèves considèrent les épreuves orales individuelles comme stressantes ou plutôt stressantes, ils et elles ne sont plus qu'environ 60% lors d'épreuves orales par deux. Si l'on ne tient compte que du pourcentage d'élèves nettement d'accord (ou stressé-e-s) pour les items « Je me sens stressé-e lors des épreuves orales individuelles/par deux en allemand », la différence est encore plus marquée : alors que près de 41% des élèves sont clairement stressé-e-s lors d'épreuves individuelles, ce pourcentage est divisé par deux et tombe à 22% lors d'épreuves par deux. De même, un grand nombre d'élèves se sentent très stressé-e-s lorsqu'ils ou elles doivent passer devant toute la classe. C'est ce qu'indique le fort taux d'approbation (53%) concernant l'item « Je me sens stressé-e lors d'une épreuve orale d'allemand qui a lieu dans la salle de classe devant les autres élèves ».

Dans le questionnaire, il a également été demandé aux élèves d'indiquer leurs préférences

en matière de feedback sur les épreuves d'expression orale. Il s'avère que tous les types de feedback sont jugés intéressants. Les voici classés par ordre de préférence (du plus élevé au moins élevé) : 1) des conseils pour s'améliorer, 2) un commentaire sur ce qui a été bien et moins bien fait pendant l'épreuve, 3) une note, 4) une grille d'évaluation complétée, 5) un feedback sur la performance de l'ensemble de la classe. Les préférences des élèves ne correspondent donc pas au feedback qu'ils et elles reçoivent de leurs enseignant-e-s qui donnent toujours une note parfois accompagnée de la grille d'évaluation complétée, ou d'un feedback à toute la classe (voir chap. 4.3). Le feedback qui intéresse le plus les élèves – des conseils pour s'améliorer ou un commentaire sur ce qui a été bien ou moins bien fait – n'est pas ou très rarement donné par les enseignant-e-s interrogé-e-s.

Dans le questionnaire, les élèves ont également indiqué brièvement ce qu'ils ou elles feraient du feedback reçu. À cet égard, il convient de noter que les élèves ne reçoivent pas toutes et tous le même feedback, ni la même quantité de feedback. Leurs réponses peuvent être classées en trois catégories⁶ : premièrement, il y a des élèves qui disent utiliser le feedback pour s'améliorer ($n=121$, 57%). Chez certain-e-s, il s'agit de déclarations très vagues comme « Je m'améliore », d'autres disent vouloir améliorer leurs erreurs ou leurs faiblesses mais ne donnent aucune indication sur le travail qu'elles et ils fournissent pour y parvenir ; d'autres encore font référence à l'effort fourni, par exemple « Guardo dove devo migliorare di più e mi alleno a casa e provo a migliorare ». À cela s'ajoutent quelques élèves, pour qui le feedback sert (à se préparer) à l'examen suivant, bien que là aussi, dans la grande majorité des cas, on ne sache pas clairement sur quoi ou comment ils et elles travaillent. Deuxièmement, il y a des élèves qui déclarent regarder le feedback ($n=66$, 30%) : une partie de ces élèves se concentre sur les erreurs ou les faiblesses, tandis que d'autres ne précisent pas ce à quoi ils et elles font attention. Troisièmement, il y a des élèves qui n'utilisent pas le feedback ($n=47$, 19%), certain-e-s ont répondu « rien » à la question, tandis que d'autres déclarent classer simplement le feedback ou le montrer à leurs parents. Les réponses des élèves indiquent qu'ils et elles sont nombreux à avoir conscience du potentiel du feedback pour la poursuite de l'apprentissage. Toutefois, leurs affirmations souvent très vagues laissent aussi supposer que les élèves n'assimilent guère le feedback et ne l'utilisent pas (ou ne peuvent pas l'utiliser) pour la suite de l'apprentissage, ce qui soulève aussi la question de savoir dans quelle mesure le feedback donné par les enseignant-e-s est parlant pour les élèves (voir chap. 4.3).

6 Certaines réponses ont été classées dans le premier groupe (« à utiliser pour un apprentissage ultérieur ») ainsi que dans le deuxième groupe (« regarder le feedback »).

5 Conclusion

Notre projet de recherche avait pour objectif de documenter les pratiques formatives et sommatives d'évaluation des compétences orales dans l'enseignement du DaF au niveau secondaire I. Il s'agissait en outre de recueillir les opinions qu'en avaient les enseignant-e-s et de connaître les avis des élèves sur l'évaluation sommative. Dans la première phase du projet, nous avons observé des leçons de DaF et surtout relevé des corrections spontanées mais très peu de séances d'évaluation formatives planifiées (voir Davison & Leung, 2009, pour un aperçu des types d'évaluation formative). Si l'on admet que les évaluations en classe devraient avoir pour objectif d'améliorer l'apprentissage des élèves et que l'on analyse les pratiques d'évaluation sous cet angle, nous relevons aussi n'avoir que rarement observé de feedback individuel utile à l'apprentissage ou qui mette en évidence des possibilités d'amélioration de la production orale lors des leçons (ou des épreuves sommatives). Cela dit, nous n'avons finalement eu qu'un bref aperçu du quotidien de chaque enseignant-e (deux leçons). Dans le questionnaire, les élèves indiquent pourtant souhaiter recevoir un feedback sous forme de conseils leur permettant d'améliorer leurs compétences d'expression orale (voir chap. 4.5).

La deuxième phase du projet a consisté à explorer comment les épreuves d'expression orale se déroulent au niveau secondaire I et comment les enseignant-e-s gèrent cette tâche complexe. L'examen des épreuves d'expression orale authentiques filmées dans huit classes a révélé que les enseignant-e-s se comportaient de manière très différente, en particulier durant les épreuves par deux. Alors qu'un enseignant (Tom) n'intervient quasiment jamais pendant une épreuve, même en cas de problèmes d'interaction évidents entre les candidat-e-s, une autre enseignante (Lea) donne déjà presque des cours de langue pendant l'examen et intervient régulièrement dans les épreuves en binôme pour diverses raisons : la gestion de l'examen, en cas de problèmes d'interaction, pour corriger, donner un feedback positif ou encore vérifier si les structures qui viennent d'être enseignées peuvent déjà être utilisées activement. Une analyse conversationnelle des épreuves orales par deux a montré que toutes les interventions n'avaient pas un effet positif sur l'interaction des candidat-e-s. Alors que les invitations à improviser ou la reformulation d'une question difficile à comprendre s'avèrent propices à la progression de l'interaction entre pairs, l'élicitation de mots et de structures en particulier entraîne des interruptions prolongées. L'élicitation est donc en contradiction avec l'objectif premier des épreuves orales par deux, à savoir l'évaluation de l'interaction orale. Elle empêche également que les avantages des épreuves orales en binôme, qui sont relevés dans la littérature spécialisée, comme le fait que les candidat-e-s prennent plus de responsabilités et soient moins réactif-ve-s que lors d'épreuves orales individuelles (Taylor & Wigglesworth, 2009; Rydell, 2019), se concrétisent. Les différences de comportement entre les quatre enseignant-e-s étudié-e-s lors des épreuves ont probablement des causes variées, par exemple le fait que les objectifs des examens ne sont pas les mêmes. Ainsi, Tom fait passer un test de *proficiency* indépendant du manuel pour véri-

fier les compétences d'interaction orale, tandis que les trois autres enseignantes souhaitent vérifier les objectifs d'apprentissage du chapitre du manuel, notamment l'acquisition de certains mots et structures. La vérification des compétences d'interaction n'était donc probablement pas au premier plan de ces tests d'*achievement*, ce qui en soi est contradictoire, puisqu'il s'agit bien d'une épreuve d'expression orale. D'une manière générale, il apparaît que les enseignant-e-s poursuivent des objectifs différents, voire divergents, au travers des épreuves orales en binôme : l'élicitation de structures n'a en effet plus grand-chose à voir avec la vérification des compétences d'interaction.

En fin de compte, il se peut que les différences de comportement des examinateur-trice-s soient aussi liées au fait que – contrairement aux tests standardisés – les enseignant-e-s jouent plusieurs rôles pendant les épreuves : à la fois gestionnaire de tests et évaluateur-trice, mais aussi enseignant-e des candidat-e-s, voire interlocuteur-trice (Teasdale & Leung, 2000; Koizumi, 2022). Notre analyse montre que les quatre enseignant-e-s pondèrent leur rôle différemment : alors que Tom est essentiellement dans le rôle de l'évaluateur et du gestionnaire de test, Lea est souvent dans le rôle de l'enseignante qui élicite les structures et les corrige de temps à autre. Comme Lea, Eva et Mia recourent à des activités d'examen qui ne répondent pas aux critères de qualité des tâches communicatives, elles doivent aussi faire beaucoup plus de gestion d'examen (par ex. organiser le tour de parole). Il ressort donc aussi de nos résultats qu'il faut sensibiliser les enseignant-e-s à leurs différents rôles pendant une épreuve et les informer du fait que ces rôles ne peuvent et ne doivent pas tous être remplis en même temps et avec la même efficacité. Alors que les épreuves orales individuelles mettent l'accent sur le rôle d'interlocuteur-trice (et non sur celui d'évaluateur-trice), les épreuves orales par deux doivent mettre l'accent sur la gestion de l'examen (et sur l'évaluation, le cas échéant).

Bien que l'évaluation des compétences ne soit pas au centre de notre étude, les sondages ont clairement révélé les critiques des enseignant-e-s à l'égard des critères d'évaluation utilisés, alors même qu'ils et elles sont tenu-e-s d'utiliser une grille d'évaluation spécifique – dans le cas de Fribourg. Cette grille est un instrument censé servir à la fois pour les épreuves orales par deux aussi bien que pour les épreuves orales individuelles (par ex. les exposés) et qui est critiqué, sans doute à juste titre, pour son caractère peu concret et différencié notamment. La critique portant sur le caractère subjectif des évaluations réalisées à l'aide de cette grille semble elle aussi justifiée : comme la grille n'est pas liée aux niveaux du CECR et qu'il n'existe pas d'exemples de référence, les évaluations reposent inévitablement sur l'expérience ou une norme subjective.

Dans l'ensemble, on peut retenir que l'adéquation entre les objectifs d'apprentissage, le but de l'examen, les tâches d'examen, le comportement de l'enseignant-e et les critères d'évaluation devrait encore être améliorée dans les épreuves que nous avons observées, afin d'assurer une évaluation aussi équitable que possible. Mais en même temps, les enseignant-e-s sont soumis-e-s à des conditions-cadres (comme des contraintes de temps et de personnel et, selon les cantons, des directives et des interdictions) qui ne simplifient pas forcément leur tâche complexe. Afin d'aider au mieux les enseignant-e-s à évaluer les com-

pétences d'expression orale, il nous semble important que, en particulier dans le canton de Fribourg, les enseignant·e·s bénéficient d'une plus grande marge de manœuvre afin de pouvoir décider librement de la meilleure manière de vérifier l'acquisition des objectifs d'apprentissage. Cela implique le choix d'une grille d'évaluation adaptée à la nature et à l'objet de l'examen, ainsi que la possibilité d'effectuer occasionnellement des tests de vocabulaire et de grammaire. Il serait ainsi possible d'éviter que les enseignant·e·s adoptent un comportement interventionniste durant une épreuve orale dans le seul but d'élucider des structures dont l'utilisation n'est pas du tout nécessaire à la réalisation de la tâche en termes de communication. Selon leurs propres déclarations, les enseignant·e·s ne manquent pas de matériel d'examen, mais notre étude met en évidence un besoin de formation initiale et continue dans différents domaines et notamment: comment distinguer entre les objectifs des épreuves, comment développer les tâches, mais aussi comment fournir des feedbacks favorisant l'apprentissage ou quel comportement adopter en tant qu'examinateur·trice·s. Afin de sensibiliser les enseignant·e·s à l'impact de leurs différents comportements durant les épreuves, nos vidéos d'examen préparées dans le cadre du projet peuvent aussi être utiles et sont à disposition sur demande.

Certaines épreuves orales filmées ont été préparées à des fins de formation initiale et continue.

Les personnes travaillant dans une haute école peuvent s'adresser à Moritz Sommet (moritz.sommet@unifr.ch), responsable du centre de documentation du Centre scientifique de compétence sur le plurilinguisme (CSP). Le document d'accompagnement (comprenant les transcriptions, les informations sur les tâches d'examen et le classement des élèves par les expert·e·s) est disponible à l'adresse suivante (en allemand): <https://centre-plurilinguisme.ch/fr/recherche/evaluation-production-orale>.

Valutazione della produzione orale nelle classi di tedesco come lingua straniera:

pratiche e opinioni di insegnanti e allieve/i di scuola media

—

Elisabeth Peyer, Gabriela Lüthi, Nadia Ravazzini

1 Situazione di partenza

Le valutazioni delle prestazioni orali – che si tratti di reagire agli interventi delle alunne e degli alunni, di dare un feedback su presentazioni o di valutare esami – appartengono alla quotidianità didattica e hanno un ruolo centrale nel processo di insegnamento e apprendimento: consentono alle allieve e agli allievi, così come alle e agli insegnanti, di rendersi conto di quanto imparato e di quali lacune rimangono; al corpo docente forniscono informazioni su come aiutare ancora meglio allieve e allievi a raggiungere i loro obiettivi. Da diversi sondaggi condotti tra insegnanti di lingue straniere è tuttavia emerso che le e i docenti tendono ad avere dubbi su come gestire le valutazioni e auspicerebbero maggiori conoscenze e nozioni pratiche, in particolare per valutare l'espressione orale (tra gli altri Berry et al., 2019; Vogt, 2018). Il presente studio si concentra dunque sulla cosiddetta *Language Assessment Literacy* delle e degli insegnanti per quanto concerne la produzione orale. Nel contesto indagato – lezioni di tedesco come lingua straniera al livello secondario I nella Svizzera francese e italiana – la *Language Assessment Literacy* è di particolare importanza perché il tedesco è obbligatorio e la nota di questa materia può essere determinante per il prosieguo della carriera scolastica delle allieve e degli allievi. Il concetto di *Language Assessment Literacy* comprende, oltre al fatto di sapere come preparare e svolgere le valutazioni delle prestazioni, anche la consapevolezza delle ripercussioni che tali valutazioni possono avere sull'individuo e sulla società (tra gli altri Inbar-Lourie, 2008; Fulcher, 2012; Vogt & Tsagari, 2014). Inbar-Lourie, per esempio, sottolinea che nella (*Language Assessment Literacy*) rientra anche la capacità di porsi domande critiche sullo scopo della valutazione, sull'idoneità dello strumento utilizzato e sulle condizioni di test, e di sapere che cosa accadrà con i risultati ottenuti dai test (2008: 389).

La verifica delle competenze orali comporta diverse difficoltà. La fugacità della comunicazione orale ne rende complicata la valutazione e il processo può essere influenzato da vari fattori che non dovrebbero tangere la valutazione delle prestazioni (varianza irrilevante al costrutto), come compiti, griglia dei criteri, valutatrice/tore e interlocutrice/tore (Isaacs, 2016; De Jong, 2023). Secondo Csépes & Fekete, per una corretta valutazione della produzione orale le e gli insegnanti devono quindi capire in particolare i principi di base dello sviluppo dei compiti, del processo di valutazione e delle griglie dei criteri (2018: 99). Per una valutazione orientata ai criteri con l'ausilio di una tabella servono inoltre una fase di esercitazione ed esempi concreti (ibidem). La valutazione sommativa della produzione orale in aula costituisce una sfida perché è onerosa in termini di tempo e rappresenta un compito molto complesso per l'insegnante, che deve assumere contemporaneamente ruoli che, per esempio nel quadro degli esami standardizzati internazionali, sono ripartiti tra più persone, come la/il responsabile del test, la/il valutatrice/tore e, secondo il formato, l'interlocutrice/tore (Teasdale & Leung, 2000; Koizumi, 2022). Tutto questo mentre continua a fare l'insegnante per le allieve e gli allievi in questione.

2 Domande di ricerca

Partendo dai presupposti summenzionati, il progetto *Valutazione della produzione orale nelle lezioni di tedesco come lingua straniera* ha indagato le pratiche di valutazione formativa e sommativa utilizzate dalle e dagli insegnanti di tedesco come lingua straniera al livello secondario I in Svizzera, e rilevato le loro opinioni al riguardo. Se in passato sono già stati svolti numerosi sondaggi sulle pratiche di valutazione, la formazione e le esigenze di perfezionamento nel settore della valutazione (Berry et al., 2019; Fulcher, 2012; Tsagari & Vogt, 2017; Vogt & Tsagari, 2014; Vogt, 2018), nell'ambito della scuola dell'obbligo vi sono ancora pochissimi studi che osservano le pratiche di valutazione e le mettono in relazione con le opinioni delle e dei docenti. Ciò sorprende, considerato che tali pratiche non possono essere rilevate direttamente e dipendono dalle esperienze e dalle predisposizioni delle e degli insegnanti (Berry et al., 2019; Scarino, 2013). Dato che anche le opinioni e i desideri delle allieve e degli allievi sono finora stati a malapena oggetto di ricerca, il nostro studio mirava a scoprire come insegnanti, alunne e alunni possano essere sostenuti al meglio per quanto riguarda la valutazione della produzione orale.

Sono state poste le domande seguenti:

1. Quali pratiche di valutazione formative adottano le e gli insegnanti di tedesco come lingua straniera del livello secondario I per quanto riguarda la produzione orale e come motivano questa scelta?
2. Quale utilità vedono le e gli insegnanti nella valutazione sommativa della produzione orale e quali esigenze (mezzi ausiliari, materiale) hanno?
3. Come vengono svolte le valutazioni sommative della produzione orale?
 - a. Quando e come intervengono le e gli insegnanti durante le prove a coppie?
 - b. Quali effetti hanno gli interventi delle e degli insegnanti sul prosieguo dell'interazione tra pari?
4. Quali opinioni e desideri esprimono le allieve e gli allievi in merito alla valutazione della produzione orale?

Alla domanda 1 si è cercata una risposta attraverso l'osservazione di lezioni e colloqui retrospettivi con insegnanti. Anche per la domanda 2 si è proceduto intervistando le e gli insegnanti. Alla domanda 3 si è risposto analizzando registrazioni video di esami, alla domanda 4 con un questionario e colloqui retrospettivi con le allieve e gli allievi filmati.

3 Fasi del progetto e metodi di ricerca

La prima fase del progetto era incentrata sulla valutazione formativa. Per rilevare le pratiche utilizzate dalle e dagli insegnanti, nonché le loro opinioni in merito, in dodici classi del livello secondario I sono state osservate due lezioni di tedesco come lingua straniera. In seguito, le e i docenti sono stati invitati a un'intervista.¹ Allo studio hanno partecipato otto insegnanti con classi (con esigenze di base e medie) dell'ottavo e del nono anno del Canton Friburgo (10H e 11H secondo HarmoS) e quattro classi del Canton Neuchâtel.² Le e i docenti disponevano di una formazione per l'insegnamento del tedesco come lingua straniera al livello secondario I, con esperienze tra i due e i venticinque anni, con una media di 7,75 anni. Hanno partecipato allo studio volontariamente. È dunque lecito supporre che si tratti di docenti con un interesse per il tema della valutazione (formativa) della produzione orale e con una certa dimestichezza in tale ambito. Visto che l'obiettivo era osservare lezioni quanto più possibile autentiche, le e gli insegnanti sono stati semplicemente invitati a tenere lezioni incentrate sulla produzione orale. I colloqui, parzialmente standardizzati, sono stati condotti al termine della seconda lezione. Suddivisi in due parti, sono durati circa un'ora. Lo scopo era rilevare l'opinione dell'insegnante sulla valutazione delle competenze orali. Nella prima parte, l'attenzione era focalizzata su circa tre momenti della lezione in cui sono state osservate pratiche di valutazione: dopo che l'intervistatrice ha menzionato una particolare situazione, l'insegnante ha spiegato per lo più direttamente come mai ha agito in quel modo e in quali situazioni adotterebbe la pratica in questione. Nella seconda parte, sono state rilevate altre pratiche e opinioni dell'insegnante sulla valutazione (formativa e sommativa) della produzione orale. I colloqui sono stati trascritti con il sistema GAT 2 (Selting et al., 2009) e in seguito valutati dal punto di vista analitico-contenutistico (Mayring, 2010).

La seconda fase del progetto era dedicata agli esami sommativi. In otto classi,³ sono state filmate prove autentiche e condotti colloqui retrospettivi con le e gli insegnanti, le allieve e gli allievi.

- 1 Desideriamo cogliere l'occasione per ringraziare le e gli insegnanti, nonché le allieve e gli allievi che hanno partecipato allo studio per la loro fiducia e la loro sincerità. La nostra riconoscenza va anche a Thomas Studer e al nostro gruppo di accompagnamento scientifico per i feedback costruttivi sulle varie tappe del progetto.
- 2 Nel Canton Neuchâtel, al settimo anno (9H secondo HarmoS) in tedesco non c'è una suddivisione in livelli. Nell'ottavo e nel nono anno (10H e 11H), la materia viene poi insegnata in due livelli. Nel Canton Neuchâtel, sono state osservate una classe del settimo anno, due classi con esigenze di base e una classe con esigenze estese
- 3 Nel Canton Friburgo hanno partecipato quattro classi: tre classi del livello delle esigenze di base (due dell'ottavo anno (10H) e una del nono (11H)), nonché una classe all'ottavo anno (10H) del livello medio. Nel Canton Neuchâtel hanno partecipato due classi: una classe del settimo anno (9H, senza livelli) e una del nono anno (11H) del livello più basso. Nel Canton Ticino hanno partecipato due classi: una classe del terzo anno di scuola media (10H) del livello più alto e una del quarto anno di scuola media (11H) del livello più basso.
- 4 Gli strumenti di rilevamento dello studio sono consultabili all'indirizzo: <https://osf.io/vwuf2/>

Le alunne e gli alunni hanno inoltre compilato un questionario sulle prove orali di tedesco come lingua straniera (cfr. la figura 1 per i dettagli sugli strumenti di rilevamento). Tutte le prove sono state svolte dalle e dagli insegnanti e, su nostra richiesta, si sono tenute perlopiù in un'aula separata. La videocamera è stata posizionata in modo da mostrare allieve, allievi e docenti. Durante le prove, il team di ricerca non era presente nell'aula perché parallelamente si è occupato del questionario scritto e dei colloqui retrospettivi con le alunne e gli alunni. I colloqui con le e gli insegnanti sono stati condotti più tardi lo stesso giorno oppure il giorno successivo online tramite MS Teams.⁴

	Metodo	Partecipanti
Fase 1 del progetto Marzo-giugno 2021	<i>Osservazione delle lezioni</i> Verbalizzazione della valutazione formativa, in particolare feedback delle e degli insegnanti	12 classi/insegnanti, 2 lezioni ciascuna (8 Friburgo, 4 Neuchâtel)
	<i>Interviste retrospettive con insegnanti</i> Parte 1 (retrospettiva) su situazioni osservate in precedenza, parte 2 su questioni generali inerenti alla valutazione	12 insegnanti
Fase 2 del progetto Marzo-giugno 2022	<i>Registrazione video di prove orali autentiche</i> Prove preparate e svolte dalle e dagli insegnanti (esami a coppie e individuali)	82 allieve/i: 48 prove in 8 classi (4 Friburgo, 2 Neuchâtel, 2 Ticino)
	<i>Interviste retrospettive con insegnanti</i> (Temi: apprezzamento della prova/della prestazione della classe, preparazione, valutazione/assegnazione delle note, utilità delle prove orali)	7 insegnanti (4 Friburgo, 2 Neuchâtel, 1 Ticino con 2 classi)
	<i>Interviste retrospettive con allieve/i</i> Brevi colloqui con le allieve e gli allievi che erano stati filmati (temi: autovalutazione, preparazione, utilità)	Interviste con 84 allieve/i
	<i>Questionario allieve/i</i> (Temi: preparazione della prova, paura della prova, feedback, utilità di vari formati di prova, autovalutazione)	254 allieve/i (17 classi) (Friburgo: 57%, Neuchâtel: 13%, Ticino: 30%)

Figura 1: panoramica dei rilevamenti dei dati

Concluso il rilevamento dei dati, le prove e i colloqui sono stati trascritti con il sistema GAT 2. Sono stati descritti anche la gestualità e la mimica. L'analisi successiva si è concentrata sulle modalità di intervento delle e degli insegnanti durante le prove a coppie, un punto sul quale è stata constatata una notevole variabilità (per maggiori dettagli, si veda Peyer et al., inoltrato per la stampa). Ne sono emersi gli interrogativi seguenti:

1. Quando e come intervengono le e gli insegnanti del livello secondario I durante le prove a coppie?
2. Quali effetti hanno i diversi tipi di intervento sull'interazione tra pari?

Per rispondere alla seconda domanda, ci basiamo sul concetto di *progressivity* (Schegloff, 2007), ossia l'avanzamento del colloquio, e partiamo dai presupposti seguenti:

- a. gli interventi delle e degli insegnanti favoriscono l'avanzamento dell'interazione tra pari se quest'ultima prosegue senza un ulteriore intervento verbale della o del docente;
- b. gli interventi delle e degli insegnanti portano poco beneficio se sono necessari ulteriori interventi verbali della o del docente prima che l'interazione tra pari possa proseguire oppure se i candidati e le candidate devono chiedere chiarimenti.

I colloqui con le e gli insegnanti, le allieve e gli allievi sono stati valutati dal punto di vista analitico-contenutistico (Mayring, 2010). Il questionario per le alunne e gli alunni comprendeva domande chiuse e aperte. Quelle aperte sono state a loro volta valutate dal punto di vista analitico-contenutistico (per maggiori dettagli sul questionario, si veda Lüthi et al., inoltrato per la stampa).

4 Risultati

Di seguito, riepiloghiamo i risultati più importanti delle due fasi del progetto.

4.1 Quali pratiche di valutazione formative adottano le e gli insegnanti del livello secondario I e come motivano questa scelta?

Nelle ventiquattro lezioni osservate, le e gli insegnanti hanno optato prevalentemente per attività strutturate (dialoghi guidati, rispondere a domande, giochi), il cui obiettivo era in primis esercitare nuovi vocaboli e strutture grammaticali. In questi casi, le e i docenti fornivano soprattutto un feedback correttivo spontaneo. Valutazioni e feedback di carattere formativo più pianificati e formali sono stati osservati raramente. La forma più frequente di feedback correttivo sono stati i *recast*, ossia la ripetizione corretta da parte dell'insegnante di frasi errate (Lyster & Ranta, 1997). Stando alle interviste, i *recast* sono apprezzati perché consentono di non dover dire esplicitamente che vi è stato un errore e – così hanno affermato due insegnanti – danno la possibilità alle allieve e agli allievi di rendersene conto da soli. Anche i *prompt* (varie indicazioni che inducono all'autocorrezione, Lyster & Saito, 2010) sono stati utilizzati regolarmente, ma non frequentemente. In poche classi, durante un'attività orale sono state osservate correzioni e spiegazioni tra pari. Durante l'intervista, una docente ha dichiarato di favorire consapevolmente le correzioni tra pari perché le allieve e gli allievi hanno un altro modo di spiegare le cose. La docente in questione ha l'impressione che tali spiegazioni risultino più comprensibili alle compagne e ai compagni. Questa opinione è condivisa anche da un'altra docente. Quattro insegnanti hanno invitato la classe (oppure un'allieva o un'allievo) ad autovalutarsi (spontaneamente). Dopo la preparazione individuale di un'attività orale, una docente ha per esempio chiesto a un alunno se fosse in grado di parlare cinque minuti sul tema in questione. L'allievo ha risposto "clairement pas". L'intento dell'insegnante – stando a quanto poi emerso nell'intervista – era di rendere la classe consapevole della necessità di migliorare ancora le conoscenze lessicali.

Le e gli insegnanti hanno dichiarato che il poco tempo a disposizione e la dimensione delle classi rendono difficile fornire feedback sulla produzione orale. Inoltre, temono che le correzioni possano demotivare le allieve e gli allievi. Correzioni e feedback vengono quindi sovente indirizzati all'intera classe, non alle singole persone. Tutte le e tutti i docenti ammettono anche che, in determinate circostanze, rinunciano di proposito a correggere, il più delle volte perché ritengono che quanto espresso sia sufficientemente comprensibile. Che cosa si intenda per comprensibilità dipende dalla situazione. Alcune e alcuni docenti dopo aver constatato un errore si chiedono se una persona di madre lingua tedesca avrebbe capito comunque il messaggio e decidono di conseguenza se correggere o meno. Il secondo motivo più frequente di una mancata correzione è la motivazione, come spiega l'insegnante 07: "J'ai pas envie de les corriger pour tout en fait [...] je peux pas toujours les casser". La

motivazione è stata un tema ricorrente dei colloqui, dai quali è emersa la difficoltà di far parlare le allieve e gli allievi. Le e gli insegnanti riconducono il problema alle scarse conoscenze lessicali e alla paura di commettere errori o di fare brutta figura. Da qui la decisione di non correggere (troppo) per non generare un effetto demotivante. Le e i docenti affermano per lo più di incoraggiare le allieve e gli allievi a parlare tedesco, gli errori hanno un'importanza secondaria: “C'est plus important pour moi qu'il parle allemand que qu'il parle JUSTE” (insegnante 03). Anche elogi e complimenti vengono utilizzati in modo mirato per rafforzare la motivazione e l'autostima. Durante le lezioni sono infatti stati sentiti brevi commenti come “molto bene” (per ulteriori informazioni sulla prima fase del progetto, si veda Peyer et al., 2023).

L'intento di salvaguardare la motivazione spiega probabilmente anche la preferenza per il *recast*. Analogamente alle e agli insegnanti interpellati nel quadro di altri studi (Kamiya, 2016; Roothoof, 2014), le e i docenti della nostra indagine hanno espresso il timore che le correzioni possano generare reazioni affettive negative. Secondo la meta-analisi di Lyster & Saito (2010), tuttavia, i *recast* sono meno efficaci dei *prompt*. Sarebbe dunque auspicabile ricorrere maggiormente a quest'ultimi. Resta in ogni caso importante prestare attenzione alla motivazione delle allieve e degli allievi, e alle loro reazioni ai feedback correttivi. Come affermato anche da Roothoof (2014), stupisce che finora non siano praticamente stati svolti studi sulle reazioni affettive delle alunne e degli alunni alle correzioni, benché si tratti palesemente di una preoccupazione che influisce sul comportamento del corpo docente.

4.2 Come vengono svolte le prove a coppie al livello secondario I?

Come menzionato in precedenza, in otto classi sono stati filmati test orali autentici. Si trattava di prove a coppie (quattro classi), di prove individuali (una classe) o di prove suddivise in due parti, una a coppie e l'altra individuale (tre classi). Di seguito, analizzeremo più in dettaglio il comportamento delle e dei quattro insegnanti che hanno svolto le prove a coppie ($n=21$), particolarmente cruciale ai livelli più bassi (A1-A2), che corrispondono a quelli della maggior parte delle allieve e degli allievi filmati. Secondo il Quadro comune europeo di riferimento per la conoscenza delle lingue (QCER), le alunne e gli alunni al livello A2 non sono infatti ancora in grado di sostenere autonomamente una conversazione (Consiglio d'Europa, 2001: 79). Secondo Galaczi, le prove a coppie al livello A2 esigono un sostegno e un controllo regolari da parte dell'insegnante (Galaczi, 2014: 557).

Attività e svolgimento delle prove

Le attività di verifica sono state elaborate dalle e dagli insegnanti, per lo più insieme a colleghe e colleghi della loro scuola. La figura 2 fornisce una panoramica delle e dei docenti, dei livelli delle classi e del tipo di prova filmato.

Insegnante	Eva ⁵	Mia	Lea	Tom
Esperienza nell'insegnamento	4 anni	25 anni	7 anni	20 anni
Classe	Classe con esigenze di base 8° anno	Classe con esigenze di base 8° anno	Classe con esigenze medie 8° anno	Classe con esigenze di base 9° anno
Attività di valutazione	Porre 5 domande sul tema della moda e rispondere		Porre 5 domande sul tema della paghetta e rispondere basandosi su una scheda	Organizzare insieme un'attività (festa, gita)
Scopo della prova	Verifica degli obiettivi di apprendimento del capitolo			Preparazione all'esame cantonale

Figura 2: indicazioni sulle e sui docenti, sulle classi e sul tipo di prova

I test di Eva, Mia e Lea servivano a verificare il raggiungimento degli obiettivi di apprendimento riportati in un capitolo del manuale didattico (*achievement test*), quello di Tom a prepararsi all'esame cantonale (*proficiency test*). Nei quattro casi, le attività di valutazione (incluse le istruzioni) erano redatte in tedesco. Dato che insegnano nella stessa scuola, Eva e Mia hanno preparato la loro attività insieme. Nel loro test, le allieve e gli allievi dovevano parlare di moda, formulando almeno cinque domande e cinque risposte a testa. Le schede riportavano diversi pronomi interrogativi e alcuni verbi in tedesco a mo' di aiuto. La prova di Lea invitava invece le allieve e gli allievi a parlare di paghetta, anche in questo caso ponendo cinque domande sul tema e rispondendovi. A questo scopo, hanno ricevuto una scheda con informazioni (per lo più sotto forma di immagini) di un/a giovane: quanto riceve, che lavoro fa, quanto desidera risparmiare. La dispensa riportava anche termini ed espressioni come “verdienen” [guadagnare], “arbeiten” [lavorare] e “genug/zu wenig” [abbastanza/troppo poco]. Nella prova di Tom, infine, le allieve e gli allievi dovevano organizzare a coppie un'attività (uscita al cinema, in piscina ecc.) concordando dettagli come gli orari, i mezzi di trasporto ecc. Anche la sua dispensa conteneva alcune espressioni in tedesco. I quattro insegnanti hanno concesso alle allieve e agli allievi tempo per prepararsi, durante il quale però non potevano prendere appunti. Le prove filmate sono durate tra i tre e i sei minuti.

5 I nomi di persone e dei luoghi menzionati nelle trascrizioni sono stati pseudonimizzati.

01	A8:	hm und in was jobbst du ((guarda A7)) hm e in cosa lavori
02	A7:	ich jobbe in äh (1.9) äh: <ah (non) ça je sais pas dire> (2.7) ah (1.0) faccio eh eh: <ah (no) questo non lo so dire>
03	Lea:	du musst nicht GENAU sagen was auf dem foto ist du kannst a:uch non devi dire ESATTAMENTE quello che c'è sulla foto puoi anche
04		((movimento circolare con la mano)) (---) einfach eine information geben fornire semplicemente un'informazione
05	A7:	okay ich jobben in (1.5) äh ich <help> ok io lavorare in eh io <aiuto>
06		((guarda Lea, si gratta la testa)) [mein jobben ist zu helfen] (-) äh [il mio lavoro è aiutare] eh
07	Lea:	[((guarda A7, annuisce più volte))]
08	A7:	neuest (-) äh: (----) neuest schu:\ (-) sch\ sch: öh schule più nuovo eh: più nuovo scuo:\ scu\ scu: eh scuola
09	Lea:	((annuisce)) hm_hm
10	A7:	öh in der <co> <of> fribourg (-) mit (-) den mathematik und französisch eh nella <co> <di> friburgo con il matematica e francese
11	Lea:	sehr gut (-) fantastisch molto bene fantastico

Estratto 2: invito a improvvisare (Lea, prova 4)

Come mostra l'estratto 2, dopo l'invito a improvvisare, A7 riesce a dire con parole semplici, e senza ulteriori aiuti, che dà ripetizioni in matematica e in francese in una scuola media, seppur zoppicando e ricorrendo occasionalmente all'inglese e al francese ("help" invece di *helfen* [aiutare] alle righe 5 e 6; "of" e "CO" (*cycle d'orientation*), ossia scuola media alla riga 10). In questo caso, quindi, basta che l'insegnante annuisca mentre l'allievo cerca le parole. Un invito a improvvisare consente pertanto di verificare che cosa le candidate e i candidati sono in grado di esprimere con mezzi linguistici molto limitati, aiutandoli a fornire un risultato che può dare un senso di soddisfazione. Inoltre, l'estratto 2 evidenzia che le aspettative concernenti lo svolgimento dell'attività sono poco chiare: permettendo all'allievo di dire quello che desidera invece di produrre quanto richiesto, Lea sostanzialmente gli semplifica l'attività. Questo aiuto dovrebbe essere considerato nella valutazione.

Altri interventi che talvolta consentono a un'interazione tra pari di avanzare sono gli inviti a chiarire nella lingua scolastica, una strategia che Mia adotta regolarmente.

01	A3:	ö:m (5.8) ä:m (1.2) wo ist (1.1) äh die shoppen ehm ehm dov'è eh il fare shopping
02		(0.9)
03	Mia:	((guarda A3))
04	A3:	((contatto visivo con Mia)) äm (1.5) Ehm
05	Mia:	<hm qu'est-ce que t'aimerais dire [S3]> <hm che cosa vorresti dire [A3]>
06	A3:	[äh] [eh]
07	A3:	<où il va faire shop\ le shopping> <dove va a fare shop\ shopping>
08	Mia:	oh W0 shoppst du ((guarda A4)) [wo gehst du] shoppen oh DOVE fai shopping [dove vai a fare] shopping
09	A3:	[ah wo shopps\ [ah dove shopp\]
10	A4:	ä:m (2.3) ich shoppe ((guarda Mia, che guarda la griglia, 1.0 sec.)) ehm faccio shopping
11		((guarda nel vuoto, 0.8 sec.)) äh: ((fa un respiro profondo)) (-) in fribourg eh: a friburgo

Estratto 3: richiesta di chiarimento nella lingua scolastica (Mia, prova 4)

Nell'estratto 3, dopo una lunga esitazione l'allievo A3 formula una domanda errata e difficile da capire con il pronome "dove" (riga 1). Mia lo guarda (riga 3) e, dopo una breve pausa, passa al francese chiedendogli quale domanda volesse porre, una domanda alla quale A3 risponde in francese. A quel punto, Mia formula una domanda corretta in tedesco (riga 8), che A3 alla riga 9 incomincia subito a ripetere. L'intervento di Mia porta dunque a un parziale "learner uptake" (Lyster & Ranta, 1997: 49). Di conseguenza, l'allievo A4 può fornire, un po' esitante, una risposta corretta senza ricorrere ad alcun aiuto. Mia pone regolarmente domande di chiarimento, benché solo nella metà dei casi circa ciò consenta di risolvere (rapidamente) un problema di interazione. Come dimostra l'estratto 3, inoltre, non basta chiedere all'allieva o all'allievo nella lingua scolastica che cosa volesse dire: in un passo successivo, l'insegnante deve tradurre la risposta in tedesco. Queste domande di chiarimento sono un segnale che le allieve e gli allievi non dispongono ancora delle strategie necessarie per risolvere autonomamente le loro difficoltà di interazione, cosa che ci si potrebbe aspettare da alunne e alunni del livello A2 (Consiglio d'Europa, 2018: 84). In generale, le allieve e gli allievi osservati paiono disporre di poche strategie, in particolare di poche strategie di compensazione per la ricerca di vocaboli.

Secondo le nostre osservazioni, alcuni interventi non contribuiscono all'avanzamento dell'interazione tra pari. È il caso in particolare dell'elicitazione di vocaboli e strutture. Nell'estratto 4, l'allievo A7 deve spiegare per che cosa spende la sua paghetta utilizzando le informazioni nella scheda, in cui sono raffigurati un regalo, patatine fritte e una salsiccia.

- 01 A7: (--) öm (-) ich kaufen mein gelb für geschenke un öh geschenke für mein freundIN
ehm io comprare la mia paghetta per regali e eh regali per mio amica
- 02 Lea: ((sorridente e annuisce))
- 03 A7: (-) und öh (1.7) äh: (1.2) <saucisse> ((guarda Lea))
e eh eh: <salsiccia>
- 04 Lea: ((guarda A7, 1.0 sec.)) was ist das wort auf deutsch für <saucisse>
come si dice in tedesco <salsiccia>
- 05 ((gesto con il palmo della mano verso l'alto))
- 06 A7: ah <mais je [sais]>
ah <ma [so]>
- 07 Lea: [oder kennst] du [öh]
[oppure sai] [eh]
- 08 A7: [[<c'est>]] öh frül frül früst ((guarda Lea))
[[<è>]] eh frül frül früst
- 09 <non> (--) ((guarda la scheda))
<no>
- 10 Lea: po
- 11 A7: ((guarda Lea))
- 12 Lea: (---) pom
- 13 A7: (-) pom f[rüst]
- 14 Lea : [pommes]
[patatine]
- 15 A7: pommes
patatine
- 16 Lea : pommes ((annuisce)) okay ((guarda A8))
patatine ok

Estratto 4: elicitare un vocabolo (Lea, prova 4)

Dato che A7 non ricorda il termine tedesco *Wurst* [salsiccia], utilizza il vocabolo francese “saucisse” e guarda l’insegnante in cerca di aiuto (riga 3). Lea allora gli chiede come si dice “saucisse” in tedesco (riga 4). Visto che A7 non riesce a rispondere (subito), la docente prova a fargli dire patatine fritte, che figurano a loro volta sulla scheda, chiedendogli “oppure sai” (riga 7). A7 sembra tuttavia essere ancora alla ricerca della parola per *Wurst*, infatti dice: “c’est öh frül frül früst” (riga 9), che potrebbe ricordare il plurale *Würste*. Lea pare non aver sentito, perché prosegue a elicitare la parola “Pommes” [patatine fritte], di cui menziona la prima sillaba. Alla fine, serviranno tre interventi di Lea per consentire ad A7 di dire, anzi, ripetere “Pommes”. Questo esempio mostra che l’elicitazione di una parola da parte dell’insegnante non favorisce necessariamente l’avanzamento dell’interazione. Al contrario, l’intervento di Lea ha interrotto a lungo l’interazione tra i due candidati e, molto probabilmente, confuso A7.

Ancora più sovente che singoli vocaboli, Lea elicitava strutture più complesse, per esempio subordinate con *damit* [affinché]. Ciò interrompe sempre l’interazione tra pari per un lasso di tempo piuttosto lungo, visto che sono spesso necessari almeno due interventi e una correzione dell’insegnante. L’obiettivo di Lea è verosimilmente quello di verificare se le candidate e i candidati sono in grado di utilizzare una struttura appena trattata a lezione. L’intervento non mira quindi a garantire l’ulteriore interazione tra le allieve e gli allievi, e quindi non persegue lo scopo principale delle prove a coppie, ossia la verifica delle competenze interattive.

4.3 Quale utilità vedono le e gli insegnanti nella valutazione sommativa della produzione orale?

Nei colloqui retrospettivi della seconda fase del progetto, alle/ai sette insegnanti partecipanti è stato chiesto come utilizzerebbero le nozioni acquisite con la prova (“Qu’allez vous faire avec les informations que vous avez obtenues grâce à cette épreuve?”). La maggior parte delle interpellate e degli interpellati ha risposto indicando quale feedback darebbe alle allieve e agli allievi dopo la prova. Alcune e alcuni docenti restituiscono la scheda compilata con la nota (ed eventualmente qualche commento sugli errori), ma dubitano che le allieve e gli allievi utilizzino questo feedback per il loro apprendimento successivo. Un’insegnante spiega quindi anche oralmente come si è arrivati al punteggio, rispettivamente alla nota. Altri due docenti, dubitando che le allieve e gli allievi traggano beneficio da un feedback scritto, forniscono a ognuno di loro la possibilità di ascoltare insieme la registrazione audio della prova al fine di formulare un feedback orale individuale. Un altro gruppo di docenti non fornisce un feedback scritto, ma è a disposizione per eventuali domande (sulla nota, risp. sulla prestazione individuale). Oltre al commento individuale, alcune e alcuni insegnanti affermano di fornire un feedback collettivo alla classe sugli errori più frequenti. Alcune e alcuni docenti dicono che con il feedback alla classe cercano di incoraggiare le allieve e gli allievi più timorosi nel tentativo di rafforzarne l’autostima. Due insegnanti aggiungono di utilizzare le informazioni acquisite per rielaborare la prova o la scheda di valutazione. Un’altro ha deciso nel corso della prova di non assegnare una nota, bensì di ripeterla dopo aver preparato meglio le allieve e gli allievi.

Alle/Ai sette insegnanti è stato chiesto anche se a loro modo di vedere le prove orali siano utili per compiere progressi in tedesco. Le risposte non sono state unanime: una metà circa ha fornito una risposta positiva, l’altra negativa. Lea, per esempio, non vede alcuna utilità: “C’est plus une source de stress ces évaluations que une source de progression”. Per le allieve e gli allievi sarebbe molto più motivante incontrare una classe della Svizzera tedesca per uno scambio. Anche un’altra insegnante esprime dubbi sull’effettivo beneficio delle prove orali: ritiene che, a quell’età, le alunne e gli alunni non siano ancora consapevoli dell’utilità di imparare il tedesco. Altre due insegnanti distinguono tra il livello dei requisiti più basso e quello più alto: alle classi più deboli che hanno partecipato allo studio le prove orali non porterebbero niente, in quanto le allieve e gli allievi in ogni caso non si pre-

parano alle prove; le classi di livello superiore, invece, studiano per la prova. I tre insegnanti restanti hanno invece un'opinione più positiva. Eva e Tom, i quali hanno partecipato allo studio con una classe con esigenze di base, vedono le prove in primis come uno strumento (di pressione) per indurre le allieve e gli allievi a lavorare in modo motivato o a esercitarsi in modo serio. Un altro insegnante argomenta diversamente: ritiene che, tra tutte le prove, quelle orali siano quelle che consentono di fare più progressi. Esse permettono alle allieve e agli allievi di rendersi conto fino a che punto sono in grado di interagire nella vita di tutti i giorni, il che li incoraggia oppure fa loro capire che devono impegnarsi maggiormente. Egli apprezza inoltre le prove orali per la possibilità di fornire alle alunne e agli alunni un aiuto più diretto rispetto a una prova scritta e un feedback immediato.

Nel complesso, tra le e gli insegnanti interpellati si constata una certa rassegnazione per quanto riguarda l'utilità delle prove orali e, in particolare, il beneficio a lungo termine dei loro feedback. Vigeva una certa perplessità su come si possa fare meglio. La maggior parte teme che le allieve e gli allievi non tengano conto dei feedback scritti (p.es. la scheda compilata), e che non li utilizzino come strumento di apprendimento. Dalla ricerca è emerso che i feedback sono propizi all'apprendimento quando sono immediati, concreti e descrittivi, e quando indicano che cosa le allieve e gli allievi possono migliorare e come farlo (Luoma, 2004; Hattie & Timperley, 2007). Un feedback (orale) di questo tipo – concreto, immediatamente successivo alla prova – è stato fornito solo da due insegnanti, verosimilmente per questioni pratiche e di tempo. Non stupisce quindi che, per vari motivi, circa la metà delle e dei sette docenti interpellate/i veda piuttosto criticamente l'utilità delle prove orali, soprattutto nelle classi con esigenze di base.

4.4 Quali esigenze hanno le e gli insegnanti per la valutazione sommativa del parlato?

Dai colloqui con le e gli insegnanti è emerso chiaramente che lo svolgimento di prove orali comporta diverse difficoltà: esse richiedono per esempio tanto tempo, in particolare quando le e i docenti procedono a registrazioni che valutano solo successivamente. Sono però state menzionate anche difficoltà di natura disciplinare che impediscono di svolgere le prove al di fuori dell'aula scolastica, in quanto la classe non può essere lasciata da sola. In tale ottica, ai dodici insegnanti della prima fase è stato chiesto se desiderassero aiuti o altro materiale per la valutazione sommativa del parlato. Le persone che insegnano a Friburgo e a Neuchâtel hanno risposto in modo sensibilmente diverso, il che è riconducibile alle rispettive condizioni generali di insegnamento. Nel Canton Friburgo, lo svolgimento di più prove orali per anno (a dipendenza delle ore settimanali) è obbligatorio, così come l'utilizzo di una determinata scheda di valutazione che distingue quattro aspetti (compito, comunicazione, correttezza formale e pronuncia/intonazione). Nel Canton Neuchâtel, invece, le prove orali vengono soltanto raccomandate, senza però fornire indicazioni sulla frequenza o sul tipo di scheda di valutazione. Alla domanda sugli aiuti desiderati, le/i quattro insegnanti di Neuchâtel hanno risposto che l'alta scuola pedagogica mette a loro disposizione materiali per

la valutazione, ma che accetterebbero volentieri altro materiale specifico che si possa usare direttamente per valutare la produzione orale, magari anche adattandolo ai diversi gruppi.

Rispondendo alla stessa domanda, le/gli otto docenti di Friburgo hanno spontaneamente menzionato la scheda di valutazione obbligatoria. La metà la critica in modo chiaro, due docenti ne criticano alcuni aspetti e due la ritengono soddisfacente. Chi esprime critiche concorda sul fatto che la scheda non consenta una sufficiente differenziazione e che vada affinata. Alcune/i segnalano la difficoltà di interpretare le categorie: "erfüllt/nicht erfüllt beim Raster – was heisst das konkret?" (Eva), rispettivamente la soggettività della valutazione: "Des fois j'ai l'impression que c'est une appréciation très personnelle finalement" (insegnante 06) o "On évalue théoriquement tous de la même façon. Voilà je dis bien théoriquement parce que je suis pas tout à fait convaincue de ça" (insegnante 04). Alcune/i criticano direttamente il principio della scheda. Lea, per esempio, trova che la valutazione sia troppo orientata alle lacune:

J'aime pas vraiment travailler avec ces grilles cantonales. Je trouve que je considère pas assez l'élève et que je me concentre sur ces quatre critères et que chaque fois qu'il y a un truc qui coince je me dis ah ben là j'enlève un point. Puis c'est vraiment j'ai l'impression une évaluation vers le bas : plutôt que de valoriser ce qu'ils ont déjà su [...]. J'ai pas l'impression que c'est le meilleur outil qu'on ait pour évaluer l'oralité.

Le e gli insegnanti del Canton Friburgo in genere auspicerebbero che la scheda venisse affinata e resa più pratica, che contenesse esempi di valutazione o che non fosse più obbligatoria. Per il resto, semmai, desidererebbero materiale pratico e differenziato per un' *esercitazione* più variata della produzione orale.

A nostro parere, le critiche alla scheda del Canton Friburgo sono giustificate: lo strumento di valutazione è per molti versi troppo poco specifico e presenta anche altre lacune. Le descrizioni dei criteri, per esempio, lasciano molto margine di interpretazione: la scheda non contiene alcun riferimento alle descrizioni delle competenze del QCER o del piano di studi né esempi concreti (registrazioni di allieve e allievi commentate e classificate). Senza esempi di riferimento, le categorie per il criterio "pronuncia e intonazione" ("Per lo più adeguate, sufficientemente adeguate, solo in parte adeguate, poco adeguate, non adeguate") possono essere applicate solo arbitrariamente. Non stupisce quindi che durante i colloqui tre insegnanti hanno affermato spontaneamente che per tale criterio assegnano lo stesso numero di punti (sufficiente) a tutte le allieve e tutti gli allievi a causa della difficoltà di valutare la prestazione. Le loro incertezze sono dovute da un lato alla genericità della scheda, dall'altro al semplice fatto che le alunne e gli alunni parlano troppo poco per poter dare loro una valutazione. Un altro punto debole della scheda è l'accumulo degli aspetti più disparati nel criterio "comunicazione", il quale comprende tra le altre cose la "variazione della struttura della frase", la "variazione del lessico" e la "concretizzazione dell'intento linguistico", che comprende l'"utilizzo corretto dei vocaboli" e la "scorrevolezza del

discorso”, aspetti che per esempio nella scheda di valutazione di Lingualevel sono ripartiti tra i criteri “spettro”, “correttezza” e “flusso” (Lenz & Studer, 2007). Molte/i insegnanti hanno inoltre osservato di essere stati formati poco o per nulla sull’utilizzo della scheda, il che mette ulteriormente in dubbio l’affidabilità di questo strumento.

4.5 Che cosa pensano le allieve e gli allievi delle prove orali?

Nell’ambito di un questionario scritto ($n=254$) e di colloqui retrospettivi ($n=84$), anche le allieve e gli allievi hanno espresso le loro opinioni e i loro desideri sulle prove orali. I principali risultati vengono presentati di seguito (per maggiori informazioni si veda Lüthi et al, inoltrato per la stampa).

Nel questionario, le allieve e gli allievi hanno valutato l’utilità dei diversi formati di prova per migliorare il loro tedesco (figura 3). Ne è emerso che trovano utili diversi formati, su tutte le prove a coppie, giudicate abbastanza utili dall’80% delle interpellate e degli interpellati, ossia più utili delle prove individuali e delle presentazioni. Nei colloqui, è stato inoltre chiesto se le prove orali aiutano a compiere progressi in tedesco. Il 57% ($n=48$), rispettivamente il 23% ($n=19$) delle allieve e degli allievi le ritiene utili o per lo meno parzialmente utili, il 18% ($n=15$) no. Nella maggior parte dei casi, le alunne e gli alunni hanno motivato la loro opinione positiva dicendo che le prove orali consentono loro di migliorare le competenze orali (14 allieve/i, 17%), che in futuro nella vita (p.es. apprendistato, lavoro) o in occasione di viaggi dovranno parlare tedesco (10 allieve/i, 12%) oppure che le prove sono un’opportunità per migliorare la pronuncia (9 allieve/i, 10%). Molti argomenti vertono dunque sul fatto che le prove orali offrono la possibilità di esercitare il parlato o aspetti della produzione orale, rispettivamente che sapere esprimersi oralmente è una capacità importante e utile e, di conseguenza, i test orali sono utili. Le allieve e gli allievi che ritengono inutili le prove orali hanno espresso diversi argomenti, il più frequente dei quali è lo stress (6 allieve/i, 7%).

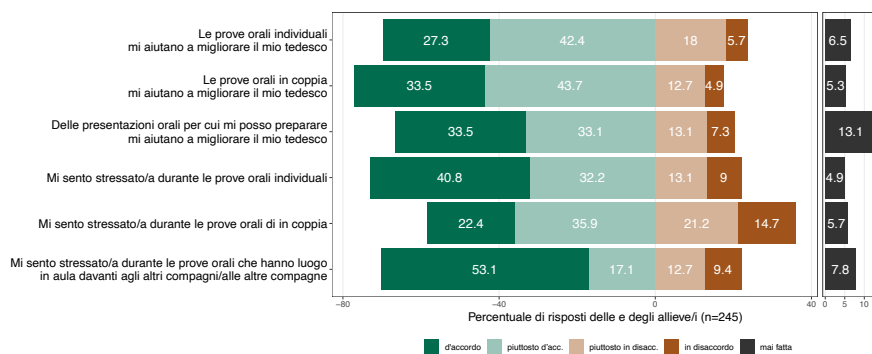


Figura 3: risultati del questionario tra le allieve e gli allievi sull’utilità e la sensazione di stress di diversi formati di prove

Nel questionario, le allieve e gli allievi dovevano indicare se si sentivano stressati con determinati formati di prova orale (figura 3). Salta all’occhio che alunne e alunni ritengono stressanti anche formati che ritengono utili. La prova a coppie è quella considerata meno stressante e al contempo più utile. Se le prove individuali sono ritenute stressanti o piuttosto stressanti da tre quarti delle interpellate e degli interpellati, per quelle a coppie la percentuale scende circa al 60%. Se per l’item “Mi sento stressato/a durante le prove orali individuali di tedesco” consideriamo solo la percentuale di coloro che si sentono molto stressati, emerge una differenza ancora più chiara: le prove individuali sono una fonte notevole di stress per quasi il 41% delle allieve e degli allievi, quelle a coppie solo per il 22%. Molte alunne e molti alunni si sentono inoltre notevolmente stressati se la prova va svolta di fronte a tutta la classe. Lo dimostra l’elevata percentuale (53%) che ha risposto positivamente all’item “Mi sento stressato/a durante una prova orale che ha luogo in aula davanti agli altri compagni/alle altre compagne”.

Il questionario chiedeva alle allieve e agli allievi di esprimersi anche sulla loro predilezione in merito ai feedback sulle prove orali. Le interpellate e gli interpellati si sono dimostrati interessati a tutti i tipi di feedback proposti. Mettendoli in ordine dal più al meno apprezzato, la sequenza è: 1) consigli per migliorare; 2) un commento su che cosa è stato fatto bene e che cosa meno durante la prova; 3) una nota; 4) una scheda di valutazione compilata; 5) un feedback sulla prestazione complessiva della classe. I desideri delle allieve e degli allievi non corrispondono dunque al feedback che ricevono dalle loro e dai loro insegnanti, i quali assegnano sempre una nota e talvolta consegnano la scheda di valutazione compilata o forniscono un feedback a tutta la classe (cfr. il punto 4.3). Il tipo di feedback al quale è interessata la maggior parte delle alunne e degli alunni – consigli per migliorare, rispettivamente un commento su che cosa è stato fatto bene e che cosa meno durante la prova – viene fornito raramente o mai.

Nel questionario le allieve e gli allievi erano chiamati a indicare brevemente anche che cosa avrebbero fatto con il feedback ricevuto. Va considerato che la qualità e il tipo di feedback ricevuti variavano da una persona all’altra. Le loro risposte possono essere suddivise in tre categorie:⁶ in primo luogo, vi sono alunne e alunni che affermano di utilizzare il feedback per migliorare ($n=121$, 57%). Alcuni danno risposte molto vaghe, come “Je m’améliore”, altri affermano di correggere errori o lacune senza però spiegare come, altri ancora forniscono maggiori dettagli, per esempio “Guardo dove devo migliorare di più e mi alleno a casa e provo a migliorare”. Vi sono poi allieve e allievi che utilizzano il feedback per prepararsi alla prova successiva, ma anche in questo caso omettono per lo più di dire a che cosa o come lavorano. Una seconda categoria è costituita da chi dice di guardare il feedback ($n=66$, 30%): alcuni si concentrano su errori o lacune, altri non descrivono nel dettaglio a che cosa prestano attenzione. La terza categoria comprende le allieve e gli allievi che non fanno uso del feedback ($n=47$, 19%). Alcuni rispondono alla domanda con “niente”, altri che si limitano ad archiviare il feedback o a mostrarlo ai genitori. Dalle risposte emerge che molte

6 Singole risposte sono state assegnate sia al primo gruppo (“Per l’apprendimento successivo”) sia al secondo (“guardare il feedback”).

alunne e molti alunni sono consapevoli dell'utilità del feedback per il loro apprendimento. I commenti spesso molto vaghi lasciano tuttavia supporre che a conti fatti i feedback non vengano praticamente utilizzati, il che solleva anche l'interrogativo di quanto le indicazioni fornite dalle e dagli insegnanti siano effettivamente utili (cfr. il punto 4.3).

5 Conclusione

Il nostro progetto di ricerca perseguiva l'obiettivo di documentare pratiche formative e sommativie per la valutazione delle competenze orali in tedesco come lingua straniera al livello secondario I, come pure di raccogliere le opinioni delle e degli insegnanti, delle allieve e degli allievi al riguardo. Durante le lezioni osservate nella prima fase sono state riscontrate soprattutto correzioni spontanee, ma poche valutazioni formative pianificate (si veda Davison & Leung, 2009, per una panoramica dei tipi di valutazione formativa). Se si analizzano le pratiche di valutazione tenendo conto del loro scopo, ossia migliorare l'apprendimento delle allieve e degli allievi, salta all'occhio che durante le lezioni (e le prove sommativie) non è stato praticamente formulato alcun feedback individuale volto a favorire l'acquisizione di competenze orali, rispettivamente a mostrare possibilità di miglioramento. Va comunque sottolineato che per ogni insegnante abbiamo assistito a sole due lezioni, insufficienti per ottenere un quadro generale di come si svolge il lavoro quotidiano in classe. Secondo il questionario, tuttavia, le allieve e gli allievi desidererebbero soprattutto ricevere consigli per migliorare la loro produzione orale (cfr. il punto 4.5).

La seconda fase del progetto doveva indagare come vengono svolte le prove orali nel livello secondario I, e come le e gli insegnanti gestiscono questo compito complesso. Dalla videoregistrazione delle prove orali autentiche di otto classi emergono grandi differenze nel comportamento delle e dei docenti, soprattutto per quanto riguarda lo svolgimento delle prove a coppie. Mentre un insegnante (Tom) non interviene quasi mai, neppure in presenza di palesi problemi di interazione delle candidate e dei candidati, un'altra (Lea) imparte quasi una lezione durante la prova e interviene regolarmente e per ogni sorta di motivo: per la gestione della prova, in caso di problemi di interazione, per correggere, per fornire un feedback positivo oppure per verificare se strutture già insegnate vengano utilizzate attivamente. Un'analisi della conversazione nelle prove orali ha rivelato che non tutti gli interventi hanno effetti positivi sull'interazione tra le candidate e i candidati. Se gli inviti a improvvisare o a riformulare una domanda difficile da capire favoriscono la prosecuzione dell'interazione tra pari, l'elicitazione di vocaboli e strutture causa invece lunghe interruzioni. L'elicitazione è dunque in contraddizione con l'obiettivo delle prove a coppie, ossia la verifica dell'interazione orale. Vengono inoltre meno i vantaggi che la letteratura specializzata attribuisce alle prove a coppie, in particolare il fatto che le candidate e i candidati assumono più responsabilità e sono meno reattivi che nelle prove individuali (Taylor & Wigglesworth, 2009; Rydell, 2019). Il diverso comportamento dei quattro insegnanti osservati può essere dovuto a vari motivi, per esempio ad uno scopo differente della prova. Tom ha svolto un test *proficiency* per verificare le competenze interattive orali, indipendente dal programma didattico, mentre le altre tre insegnanti desideravano controllare il raggiungimento degli obiettivi di apprendimento dell'unità, nei quali rientrava anche l'acquisizione di determinati vocaboli e strutture. In questi test *achievement*, quindi, la verifica delle competenze interattive non era la priorità, il che tuttavia è contraddittorio, trattandosi di una prova di

conversazione. Nel complesso, si constata che con le prove a coppie le e gli insegnanti perseguono obiettivi diversi e a tratti divergenti: l'elicitazione di strutture non ha più molto a che fare con la verifica delle competenze di interazione.

Le differenze di comportamento potrebbero essere riconducibili anche al fatto che, contrariamente a quanto avviene nei test standardizzati, in questi casi le e gli insegnanti devono assumere più ruoli: oltre a essere appunto insegnanti delle candidate e dei candidati, secondo le circostanze sono anche amministratrici e amministratori del test, valutatrici e valutatori, e interlocutrici e interlocutori (Teasdale & Leung, 2000; Koizumi, 2022). La nostra analisi rileva che le/i quattro insegnanti interpretano diversamente i loro ruoli: se Tom assume prevalentemente quello di valutatore e amministratore del test, Lea è più spesso l'insegnante che elicitava le strutture e talvolta corregge. Dato che Lea, Eva e Mia ricorrono ad attività che non soddisfano i criteri qualitativi di un compito di comunicazione, la gestione della prova risulta più impegnativa (p.es. organizzazione dei turni di parola). I nostri risultati mostrano anche che occorre sensibilizzare le e i docenti sui vari ruoli durante una prova e far loro notare che tali ruoli non possono e non devono essere svolti tutti contemporaneamente e con la medesima qualità. Se nelle prove individuali è preponderante il ruolo dell'interlocutrice/tore (e non parallelamente quello della/del valutatrice/tore), nelle prove a coppie l'attenzione deve concentrarsi sull'amministrazione (ed eventualmente sulla valutazione).

Il nostro studio non era volto a valutare le prestazioni. Ciò nonostante, dai colloqui con le e gli insegnanti è emerso chiaramente che rimettono in questione con occhio critico i criteri di valutazione utilizzati ma, nel caso del Canton Friburgo, l'impiego di una determinata scheda è vincolante. Si tratta di un unico strumento pensato per essere utilizzato sia per le prove a coppie sia per quelle individuali (p.es. presentazioni), considerato a ragione troppo poco pratico e differenziato. Anche la critica secondo la quale le valutazioni con la scheda siano soggettive pare giustificata: dato che la scheda non si riallaccia al Quadro comune europeo di riferimento per la conoscenza delle lingue (QCER) e non contiene esempi di riferimento, le valutazioni si basano forzatamente sull'esperienza o su una norma soggettiva.

Nel complesso, constatiamo che, nell'ottica di una valutazione quanto più possibile equa, la congruenza tra gli obiettivi di apprendimento, lo scopo delle prove, i compiti, il comportamento delle e degli insegnanti e i criteri di valutazione andrebbero migliorati. Al contempo, però, siamo consapevoli che le e i docenti sono vincolati a condizioni generali di insegnamento (come risorse scarse in termini di tempo e di personale, direttive e divieti a seconda del Cantone) che non agevolano di certo un compito già complesso. Al fine di sostenere le e gli insegnanti nella valutazione della produzione orale, ci pare importante concedere maggiori libertà, in particolare a quelle e quelli del Canton Friburgo, affinché possano decidere autonomamente come verificare al meglio gli obiettivi di apprendimento. In questo rientrano anche la scelta di una scheda di valutazione commisurata al tipo e allo scopo della prova, nonché la possibilità di svolgere, di tanto in tanto, test lessicali e grammaticali. Ciò consentirebbe di evitare che durante una prova orale le e gli insegnanti adottino un comportamento interventista solo per elicitarne strutture il cui utilizzo dal punto di

vista comunicativo non è necessario per il compito in questione. Secondo dichiarazioni delle e degli insegnanti, non manca materiale per le verifiche. In compenso, dal nostro studio emerge un fabbisogno di formazione e perfezionamento in vari ambiti, in particolare nella distinzione dello scopo delle prove, nell'elaborazione dei compiti, nella formulazione di feedback che favoriscano l'apprendimento e nel comportamento delle valutatrici e dei valutatori. Le videoregistrazioni che abbiamo realizzato nel quadro del progetto possono aiutare a rendere consapevoli le e i docenti degli effetti dei diversi approcci di valutazione. Su richiesta, le mettiamo volentieri a disposizione.

Alcune prove orali registrate sono state elaborate per scopi di formazione e perfezionamento.

Chi lavora in una scuola universitaria può richiederle a Moritz Sommet (moritz.sommet@unifr.ch), responsabile del centro di documentazione del Centro scientifico di competenza per il plurilinguismo (CSP). Il relativo documento di accompagnamento (con trascrizioni, informazioni sui compiti delle prove e graduatorie di esperte ed esperti delle prestazioni delle allieve e degli allievi, in tedesco) si trova su <https://centre-plurilinguisme.ch/it/ricerca/valutazione-produzione-orale>.

Assessment of speaking skills in the German as a foreign language classroom:

practices and views of lower secondary school teachers and students

—

Elisabeth Peyer, Gabriela Lüthi, Nadia Ravazzini

1 Starting point

Assessments of oral performance – whether as a reaction to student input, as feedback on presentations, or in test situations – are an inherent part of everyday school life and they play an instrumental role in teaching and learning. Assessments help both students and teachers see how well subject matter has been mastered and where gaps persist. In ideal cases, the information teachers gain from assessments also enables them to better respond to their students' needs and, in doing so, foster learning. However, various surveys of foreign language teachers have revealed that teachers often feel uncertain when evaluating their students and that they would like to have greater access to knowledge and practical tools for addressing questions surrounding assessment, including the assessment of speaking skills (Berry et al., 2019; Vogt, 2018). These are the conditions that form the backdrop of this study, which is interested in teachers' so-called language assessment literacy with regard to speaking skills. In the setting chosen for the study – German as a foreign language (GFL) lessons at the lower secondary level in French- and Italian-speaking Switzerland – a teacher's language assessment literacy is of particular importance, as German is a compulsory school subject and a student's German marks can have a decisive impact on their future school trajectory. The concept of language assessment literacy is used to describe not only the knowledge and skills needed to plan and carry out assessment activities and evaluate student performance – it also encompasses an awareness for the impacts that assessments can have on individuals and society as a whole (Inbar-Lourie, 2008; Fulcher, 2012; Vogt & Tsagari, 2014). As an example of this broader significance, Inbar-Lourie stresses that (language) assessment literacy also includes the ability to ask critical questions about the purpose of an assessment, the efficacy of a chosen instrument, and the testing conditions as well as an understanding about what future decisions will be based on the test results (2008: 389).

The assessment of speaking skills goes hand-in-hand with various challenges, including the fact that assessing a performance is complicated by the fleeting nature of spoken language. In addition, certain factors that should have no impact on an assessment result can be present in assessment processes (construct-irrelevant variance); these include test tasks, criteria rubrics, raters or even interlocutors (Isaacs, 2016; De Jong, 2023). According to Csépes & Fekete, a valid assessment of speaking skills requires that teachers must in particular have a good understanding of the basic principles of task design and of assessment procedures as well as the ability to work with assessment rubrics (2018: 99). Moreover, in the case of criteria-based assessments using a rubric, teachers should receive appropriate training and practical reference examples should be given (2018). Summative assessments of speaking skills in the classroom bring their own set of challenges due to the time-consuming nature of these activities but also because they present teachers with a particularly complex task. In particular, the setting demands that teachers simultaneously take on multiple roles – that of test administrator, rater and, depending on the test format,

interlocuter – that in standardised international tests are generally distributed across several people (Teasdale & Leung, 2000; Koizumi, 2022). And at the same time, they also retain their role as teacher of the students being tested.

2 Research questions

Against the backdrop described above, the project *Assessment of speaking skills in the German as a foreign language classroom* examined the formative and summative assessment practices that teachers of GFL at lower secondary schools in Switzerland used to evaluate their students' speaking skills; in addition, information was also gathered on teacher attitudes towards these practices. Although various teacher surveys have already been conducted on assessment practices or on training and additional professional development needs in the area of language testing and assessment (Berry et al., 2019; Fulcher, 2012; Tsagari & Vogt, 2017; Vogt & Tsagari, 2014; Vogt, 2018), only very few studies – especially in the context of compulsory schooling – have observed actual assessment practices and linked these practices to teacher views. This omission is all the more surprising, as assessment practices cannot be directly established by means of a questionnaire or interview, and because these practices are interwoven with a teacher's individual beliefs and experiences (Berry et al., 2019; Scarino, 2013). Moreover, because previous research on assessment has rarely paid attention to the views and wishes of students – who are directly impacted by the practices chosen – the study placed value on including their voices in the considerations. Indeed, the overarching aim of the project was to learn more about how to best support both teachers and students in the assessment of speaking skills.

The following research questions were explored:

1. What formative assessment practices do GFL teachers at the lower secondary level use to evaluate speaking skills and how do they explain their practices?
2. What benefits of summative assessments of speaking do the teachers identify, and what are their related needs (for support, materials)?
3. How are summative assessments of speaking skills carried out?
 - a. When and how do teachers intervene during paired speaking tests?
 - b. How do these teacher interventions impact the further progression of peer interactions?
4. What views and wishes do students have concerning the assessment of their speaking skills?

Answers to question 1 were sought using information gained from classroom observations and retrospective teacher interviews; answers to question 2 also drew on the teacher interviews. Question 3 was examined using videographed authentic speaking tests, while a student questionnaire and retrospective interviews with the filmed students formed the basis for exploring question 4.

3 Project phases and research methods

The initial project phase focused on formative assessment practices. In order to investigate the kinds of formative assessments used by teachers as well as teacher attitudes towards assessment, two GFL lessons were observed in twelve separate classes at the lower secondary level and teacher interviews were subsequently conducted.¹ Eight teachers of lower- and intermediate-track GFL classes in years 8 and 9 (10H and 11H according to the HarmoS curriculum) in the Canton of Fribourg and four classes from the Canton of Neuchâtel participated in the study.² All participating teachers hold a GFL teaching qualification for the lower secondary level and their related teaching experience ranged between 2 and 25 years, with an average of 7.75 years. Teacher participation in the study was voluntary, which suggests that these teachers were already interested in the “(formative) assessment of speaking” and that they most likely had a measure of confidence in this area. Because the study aim was to observe lessons that were as authentic as possible, the teachers were given no specific task and were simply asked to conduct lessons that “focus on speaking”. The teacher interviews were semi-structured and held after the second lesson; they consisted of two parts, lasted approximately one hour and were used to document teacher attitudes towards the assessment practices. In the first part of the interview, the teachers were asked about some three situations in which assessment practices were observed. After the interviewer gave a general description of the situation, the teachers generally explained directly why they intervened as they did and in which situations they would choose this practice. In the second part of the interview, other teacher practices and views regarding (formative and summative) assessment of speaking skills were recorded. GAT 2 minimal transcription (Selting et al., 2009) was used to transcribe the interviews, which were subsequently examined using methods of content analysis (Mayring, 2010).

In the second project phase, the focus was placed on summative speaking tests. Authentic speaking tests were filmed in eight secondary school classes;³ afterwards, retrospective interviews were held with teachers and students. In addition, the students com-

- 1 We would like to take this opportunity to thank all participating teachers and students for their trust and candour over the course of the study. We also would like to express our gratitude to Thomas Studer and our academic advisory group for their constructive feedback on the various project phases.
- 2 In the Canton of Neuchâtel, no specific academic tracks are distinguished for German lessons in year 7 (9H in the HarmoS curriculum). Starting years 8 and 9 (10H and 11H), German is taught at two skill levels. In Neuchâtel, one 7th-year class was observed in the study as were two classes with basic German requirements and one class with higher requirements.
- 3 In the Canton of Fribourg, four classes participated in the study: three classes from the lower track (two 8th-year classes [10H] and one 9th-year class [11H]) as well as one 8th-year class (10H) from the intermediate track. In the Canton of Neuchâtel, two classes participated: one 7th-year class (9H, no specific academic track) and one 9th-year class (11H) from the lower track. Two classes from the Canton of Ticino also took part: one 8th-year class (10H) from the higher track and one 9th-year class (11H) from the lower track.

pleted a questionnaire about speaking tests in GFL lessons (see figure 1 for details on the survey instruments). All speaking tests were conducted by the teachers themselves, with most tests taking place in a separate room at the request of the project team. The camera was set up to ensure that students and teachers were visible on the video. During the tests, the researchers were not physically present in the exam room; rather, they conducted the written student survey and the retrospective student interviews. The retrospective teacher interviews were carried out online via MS Teams later the same day or the next day.⁴

	Methods	Participants
Project phase 1 March–June 2021	<i>Classroom observation</i> Documenting formative assessment interventions, in particular teacher feedback.	12 classes/teachers for 2 lessons each (8 from Fribourg, 4 from Neuchâtel)
	<i>Retrospective teacher interviews</i> Part 1 on observed assessment situations (retrospective); Part 2 on general questions about assessment practices	12 teachers
Project phase 2 March–June 2022	<i>Videography of authentic speaking tests</i> Tests designed and conducted by the teacher (paired and individual speaking tests)	82 students; 48 tests in 8 classes (4 from Fribourg, 2 from Neuchâtel, 2 from Ticino)
	<i>Retrospective teacher interviews</i> (Topics: general comments on test and class performance; preparation; assessment/marks; learning value of oral exams)	7 teachers (4 from Fribourg, 2 from Neuchâtel, 1 from Ticino with 2 classes)
	<i>Retrospective student interviews</i> Short interviews with filmed students (topics: self-assessment, preparation, value of tests)	Interviews with 84 students
	<i>Student questionnaire</i> (Topics: exam preparation; exam anxiety; feedback; value of different test formats; self-assessment)	254 students (17 classes) (Fribourg: 57%, Neuchâtel: 13%, Ticino: 30%)

Figure 1: Overview of the data collection

4 The survey instruments used in the study can be accessed at <https://osf.io/vwuf2/>

The recorded speaking tests and interviews – including information on gestures and facial expressions – were later transcribed using GAT 2 minimal transcription. The subsequent conversation analysis focused on teacher intervention behaviours during the paired speaking tests, as a high degree of variability was observed in this area (for details, c.f. Peyer et al., submitted). In the process, the following key research questions emerged:

1. When and how do teachers at the lower secondary level intervene during paired speaking tests?
2. What effect do different interventions have on peer interactions?

To answer the second question, the study drew on the concept of “progressivity” (Schegloff, 2007) – how well a discussion progresses – which was operationalised as follows:

- a. Teacher interventions promote the progression of peer interactions when students are able to continue speaking after an intervention, without requiring an additional verbal intervention from the teacher.
- b. Teacher interventions tend to hinder the progression of peer interactions when additional verbal interventions become necessary to enable students to resume a peer interaction or when students need to ask a clarifying question to continue.

Content analysis (Mayring, 2010) was applied to study the teacher interviews. The student questionnaire consisted of open and closed questions; content analysis was also used to study the open questions (for details on the student surveys, c.f. Lüthi et al., submitted).

4 Findings

The key findings from both project phases are summarised below.

4.1 What formative assessment practices do lower secondary school teachers use and how do they explain their decisions?

In the 24 lessons observed, teachers mainly led structured speaking activities (working with dialogue frames, answering questions, playing games) with the primary aim of practising new vocabulary and grammatical structures. It became apparent that teachers mostly offered spontaneous corrective feedback; by contrast, planned, more formal formative assessment or feedback practices were only rarely observed. The most common type of corrective feedback took the form of recasts, in which a teacher repeats what a student said but corrects any errors (Lyster & Ranta, 1997). According to statements made in the teacher interviews, some teachers particularly like using recasts because there is no need to explicitly say a student has made an error; rather – as two teachers assume – the students would realise the error on their own. The teachers also regularly, albeit infrequently, used prompts (i.e. various comments designed to encourage self-correction, cf. Lyster & Saito, 2010). Peer correction and explanation were observed during the speaking activities in just a few classes. In one interview, a teacher said she intentionally promotes peer correction because students have a way of explaining that she believes is more accessible to other students – an opinion shared by another teacher participating in the study. It was also observed that four teachers invited an individual student or the entire class to make a (spontaneous) self-assessment. For example, after students had prepared a speaking task individually, one teacher asked a student if he could speak for five minutes on the topic. The student replied, “clairement pas” [no way]. By asking the question, the teacher – as stated later in the interview – wanted students to understand that they needed to improve their vocabulary.

During the interviews, it became evident that the teachers generally found it difficult to give feedback on speaking skills due to time constraints, but also on account of large class sizes. In addition, the teachers worried that correcting errors would discourage their students. For this reason, corrections and feedback were often addressed to the entire class and not individualised. Moreover, all interviewed teachers stated that they deliberately refrained from correcting their students in certain situations. The most common reason given for not correcting errors was when they felt a student’s production was good enough to be understood. What precisely is considered comprehensible, however, is likely to depend on the specific context. For example, several teachers said they ask themselves whether a native German speaker would understand a statement despite the fact that there was an error, and that their decision to correct or not correct depends on the answer to this question. The second most common reason for not correcting concerns student motivation.

Teacher 07, for instance, said: “J’ai pas envie de les corriger pour tout en fait [...] je peux pas toujours les casser” [I didn’t want to correct every single mistake (...) I can’t always be hounding them]. Concerns about learner motivation were a recurring theme during the interviews, with the teachers saying it was already difficult enough to get their students to say anything at all. They attribute this to poor vocabulary skills or to a fear of making errors and being exposed in the classroom. Because they worry about discouraging their students from speaking, the teachers are careful to not correct (too) much. As a rule, they stress that they want to motivate their students to speak German and that mistakes are of secondary importance: “c’est plus important pour moi qu’il parle allemand que qu’il parle JUSTE” [it’s more important to me that they speak German than that they speak it CORRECTLY] (teacher 03). Praise and compliments are used intentionally with the aim of strengthening student motivation and self-esteem. In this regard, the most common type of teacher encouragement observed during the lessons were brief comments such as “very good” (for further results of the first project phase, cf. Peyer et al., 2023).

The general aim of retaining student motivation is likely also the reason that teachers often chose the more implicit correction type of recasts. And in line with teacher statements from other studies (Kamiya, 2016; Roothoof, 2014), several teachers expressed concerns that corrections would trigger negative affective reactions in their students. However, meta-analysis of classroom studies in Lyster & Saito (2010) has demonstrated that recasts are less effective than prompts, meaning more frequent use of the latter would be desirable. To be sure, it is essential that teachers are sensitive to student motivation and students’ affective reactions to corrective feedback. As Roothoof (2014) also points out, however, the fact that very few studies have been conducted on how learners react to corrections is surprising, especially because teachers obviously make assumptions about the impact, which in turn influences how they correct their students.

4.2 How are paired speaking tests conducted at the lower secondary level?

As stated above, authentic speaking tests were filmed in eight classes. The tests were either in the form of paired speaking tests (four classes), individual tests (one class), or they consisted of two parts: a paired speaking test and an individual test (three classes). The following section takes a closer look at the examiner behaviours of four teachers who conducted paired speaking tests ($n=21$). It should be noted that teacher behaviours are considerably more relevant in the case of paired speaking tests at lower language levels (A1 to A2) – which corresponds to most of the participating classes – than at higher levels, as the Common European Framework of Reference for Languages (CEFR) guidelines specify that A2 learners are unlikely to be able to maintain a conversation without outside help (Council of Europe, 2001: 79). Galaczi also remarks that paired speaking tests at the A2 level still regularly require support and monitoring by the examiner (2014: 557).

Test tasks and procedures

The teachers designed the tests themselves, generally in collaboration with other teachers at their schools. Figure 2 provides an overview of the teachers, school years and types of test filmed.

Teacher	Eva ⁵	Mia	Lea	Tom
Teaching experience	4 years	25 years	7 years	20 years
Year	lower track Year 8	lower track Year 8	intermediate track Year 8	lower track Year 9
Assessment task	Ask and answer 5 questions about fashion		Ask and answer 5 questions about pocket money using pre-defined information	Work together to organise an activity (party, excursion)
Purpose of test	Testing learning objectives of the textbook chapter			Preparation for the cantonal final exam

Figure 2: Information about teachers, school years and test types

Eva, Mia and Lea tested their students to determine how well they had mastered the learning objectives in a textbook chapter (achievement test) whereas Tom’s students took a test in preparation for the cantonal final exam (proficiency test). All four teachers wrote the exam activity (including the instructions) in German. Eva and Mia designed their test together, as they taught at the same school. Their students were required to ask and answer at least five questions on the topic of fashion. As an aid, various interrogative pronouns and some verbs in German were included on the exam sheet. In Lea’s exam, the students were required to ask and answer five questions on the topic of pocket money. For this, they were given a description of a young person with information (generally in pictures) about how much pocket money they get, what their job is, and what their savings target is. Expressions such as “verdienen” [earn], “arbeiten” [work] and “genug/zu wenig” [enough/not enough] were also given in German. For Tom’s test, the students worked in pairs to organise an activity (e.g.

5 Names of individuals and places in the transcripts have been pseudonymised.

going to a film or the swimming pool) and talk about related topics such as time and means of transport. Tom’s exam sheet also included several useful phrases in German. All teachers gave their students time to prepare, but the students were not allowed to take notes. The tests filmed for the study lasted between three and six minutes.

When and how do teachers at the lower secondary level intervene during (paired) speaking tests?

The four teachers observed in the study displayed very different behaviours during the paired speaking tests. For example, teacher interventions were highly variable, both in terms of frequency (very rarely in Tom’s case, very often in Lea’s) and function. Whereas Tom refrained from intervening – even when students were obviously having interactional troubles – and only reminded his students to summarise their discussion at the end of the test, the other three teachers intervened not only to manage the test (e.g. inviting students to ask a question) but also to help students when interactional troubles arose (comprehension difficulties, word searches) and to give them positive (usually non-verbal) feedback. At times, Mia and Lea also intervened to correct a student, generally by asking a clarifying question or through a recast. In addition, Lea intervened several times in order to control whether the students were able to use grammatical structures (e.g. causal or finite subordinate clauses) that had recently been taught. It is likely that Lea’s frequent elicitation of grammar during the test is due to a Canton of Fribourg policy introduced a few years ago prohibiting teachers to conduct grammar and vocabulary tests in foreign language lessons; according to the policy, grammar and vocabulary should be tested when communicative language abilities are assessed (SEnOF, 2018).

On the whole, the considerable variability in teacher behaviours will have various origins, including differences in test objectives and test activities. Indeed, it is unlikely to be by chance that only Tom, who rarely intervened, had chosen an assessment activity that in many aspects corresponds to a task in the technical sense (e.g. regarding goal orientation and authenticity) and that is very similar to test tasks found in standardised proficiency exams for A2 learners. The activities chosen by the other teachers included pre-arranged turn-taking and they neither required a natural use of language nor did they promote negotiations between students – which previous research has identified as advantages of paired speaking tests (May, 2009; Rydell, 2019). Lea’s exam sheet in particular caused comprehension problems (due to a misleading image, among other things), so that she often had to answer questions about the activity. In addition, individual personality traits appear to play a key role in how and how often a teacher intervenes in paired speaking tests, as seen in the different behaviours observed in Eva and Mia, who used the same exam with students at the same language level.

What effects do different kinds of interventions have on peer interactions?

This section analyses common and repeatedly observed intervention strategies in relation to their impact on the progression of a peer interaction, beginning with interventions that promote the peer interaction, and then considering those that tend to hinder it.

One of the most effective interventions observed in the study is the reformulation of a question that is difficult to understand – a type of mediation strategy that Mia regularly applied with positive results. Excerpt 1 provides a typical example of this strategy.

- 01 S1: äm:: (6.1) wo (2.9) wo: (2.8) <non> (1.3) ((whispers incomprehensible, ca. 1 sec.);
um:: where where <no>
- 02 welche (1.3) trage ((taps her index finger on the table with every word)) (-)
which wear
- 03 schwarz (-) t-shirt ((looks at S2))
black t-shirt
- 04 (1.0)
- 05 Mia: ((looks at S2)) trägst du schwarze t-shirts
do you wear black t-shirts
- 06 S2: äh ja ich: äh: trage (-) äh schwarze t-shirt
uh yes i: uh: wear uh black t-shirt

Excerpt 1: Reformulation of an unclear question (Mia, exam 5)

In halting language interspersed with longer breaks, student S1 tries to ask a question using the pronoun “wo” [where] in line 1, but then attempts a different approach and begins to ask a second question with “welche” [which] in line 2. In the end, however, S1’s question is unclear (which wear black T-shirt?), and Mia intervenes with a direct question for student S2: “trägst du schwarze T-Shirts?” [Do you wear black T-shirts?]. S2 can respond to this question with little hesitation. It is, however, unclear whether Mia’s reformulation is what S1 actually meant to say and whether her intervention was indeed necessary – or whether the students would have been able to solve their (potential) interaction difficulties on their own. S2 could have asked a clarifying question, for example.

Another intervention strategy that consistently furthered the conversation is when a teacher invited a student to improvise, that is to not adhere strictly to the information on the exam sheet. Lea occasionally used this strategy when students were looking for words or when it became clear they were struggling to find the right expression. In excerpt 2, student S7 is having difficulties in formulating the fact that – according to the exam sheet – he has a side job as a cleaner. Lea encourages S7 to provide any information he likes (line 4).

- 01 S8: hm und in was jobbst du ((looks at S8))
hm and in what do you work
- 02 S7: ich jobbe in äh (1.9) äh: <ah (non) ça je sais pas dire> (2.7) äh (1.0)
i work in uh uh: <ah (no) I don't know how to say that> uh
- 03 Léa: du musst nicht GENAU sagen was auf dem foto ist du kannst a:uch
you don't have to say EXACTLY what is on the photo you can a:so
- 04 ((circular hand movement)) (---) einfach eine information geben
just give an information
- 05 S7: okay ich jobben in (1.5) äh ich <help>
okay i work in uh i <help>
- 06 ((looks at T03, scratches his head)) [mein jobben ist zu helfen] (-) äh
[my job is to help] uh
- 07 Léa: [((looks at S7, nods))]
- 08 S7: neuest (-) äh: (----) neuest schu:\ (-) sch\ sch: öh schule
newest uh: newest schoo:\ sch\ sch: uh school
- 09 Léa: ((looks at S7, nods)) hm_hm
- 10 S7: öh in der <co> <of> fribourg (-) mit (-) den mathematik und französisch
uh in the <co> <of> fribourg with the maths and French
- 11 Léa: sehr gut (-) fantastisch
very good fantastic

Excerpt 2: Invitation to improvise (Lea, exam 4)

After being told he could improvise, S7 uses simple words, albeit haltingly and with occasional recourse to English and French – “help” instead of *helfen* in lines 5 and 6; “of” and “CO” (*cycle d’orientation*) for “secondary school” in line 10 – but with no further assistance to say he works as a maths and French tutor at a secondary school. That the teacher nods encouragement is enough to support him while he searches for the right words. The invitation to improvise is what enables the teacher to control what students with very limited linguistic resources can express and to possibly help them attain a sense of achievement. At the same time, however, excerpt 2 also highlights that it is unclear what Lea expected of the students: Lea simplified the test by allowing the student to deviate from the prescribed content. Consequently, she would have to factor in this assistance when giving marks.

The intervention types that sometimes contribute to maintaining a peer interaction include requests for clarification in the language of schooling, a strategy that Mia regularly chose.

01 S3: ö:m (5.8) ä:m (1.2) wo ist (1.1) äh die shoppen
 u:m u:m where is uh the shopping

02 (0.9)

03 Mia: ((looks at S3))

04 S3: ((eye contact with Mia)) äm (1.5)
 um

05 Mia: <hm qu'est-ce que t'aimerais dire [S3]>
 <hm what would you like to say [S3]>

06 S3: [äh]
 [uh]

07 S3: <où il va faire le shopping>
 <where he goes shopping>

08 Mia: oh W0 shoppst du ((looks at S4)) [wo gehst du] shoppen
 oh WHERE do you shop [where do you] go shopping

09 S3: [ah wo shopp\]
 [ah where do shop]

10 S4: ä:m (2.3) ich shoppe ((looks at Mia, who is looking at rubrics, 1.0 sec.))
 u:m i shop

11 ((stares into space, 0.8 sec.)) äh: ((takes a deep breath)) (-) in fribourg
 uh: in fribourg

Excerpt 3: Clarification question in the language of schooling (Mia, exam 4)

In excerpt 3, student S3 takes a long time to formulate an incorrect and difficult to understand question with the pronoun “wo” [where] (line 1). Mia looks at him (line 3) and, after a short pause, switches to French to ask what he wants to say, and S3 answers the question in French. Mia then formulates a correct question in German (line 8) that S3 immediately begins to repeat in line 9. As such, Mia’s intervention brings about a partial “learner uptake” (Lyster & Ranta, 1997: 49). His partner, student S4, is then able to offer a hesitant answer without needing to ask for help. Mia regularly uses clarifying questions when students encounter problems, but it should be noted that these interventions lead to a (quick) resolution only half the time. Excerpt 3 also shows that simply resorting to the language of schooling to ask a student what they want to say is not enough to promote dialogue: the teacher must take an additional step and translate the student’s reply into German. On another note, these clarifying questions also suggest that students lack the necessary strategies to solve interactional troubles on their own, which could be expected at the A2 level (Council of Europe, 2018: 84). In general, the students in the study appear to have mastered very few strategies, in particular compensation strategies for when they have difficulty finding the right word.

Some interventions – as observed during the study – failed to promote the progression of peer interactions, in particular when teachers wanted to elicit vocabulary or grammatical structures. In excerpt 4, student S7 was asked to say what he spends his pocket money on and refer to the pre-defined information (images of a gift, French fries and a sausage).

01 S7: (--)öm (-) ich kaufen mein gelb für geschenke un öh geschenke für mein freundIN
 um I buy my money for presents and uh presents for my girlfriend

02 Lea: ((smiles and nods))

03 S7: (-) und öh (1.7) äh: (1.2) <saucisse> ((looks up to Lea))
 and uh uh: <sausage>

04 Lea: ((looks at S7, 1.0 sec.)) was ist das wort auf deutsch für <saucisse>
 what is the word in german for <sausage>

05 ((gesture with palms facing upwards))

06 S7: ah <mais je [sais]>
 ah <but I [know]>

07 Lea: [oder kennst] du [öh]
 [or do you know] [uh]

08 S7: [[<c'est>]] öh frül\ frül\ früst ((looks at Lea))
 [<(it's)>] uh frül frül früst

09 <non> (--) ((looks down at exam sheet))
 <no>

10 Lea: po

11 S7: ((looks at Lea))

12 Lea: (---) pom

13 S7: (-) pom f[rüst]

14 Lea : [pommes]
 [fries]

15 S7: pommes
 fries

16 Lea : pommes ((nods)) okay ((looks S8))
 fries

Excerpt 4: Eliciting a word (Lea, exam 4)

S7 is unable to find the German word *Wurst* [sausage], so he uses the French, “saucisse”, and looks to his teacher for help (line 3). Lea asks if he knows the German word for “saucisse” (line 4). When S7 is unable to (immediately) answer, she then tries to elicit the word for fries, which also are pictured, by asking, “oder kennst du” [or do you know] (line 7). However, it seems that S7 is still looking for the word *Wurst*, because he says: “c’est öh frül\ frül\ früst” [it’s uh frül\ frül\ früst] (line 9), which is similar to the plural form, *Würste*. Lea apparently fails to hear him and continues to elicit the word “Pommes” [fries] by pronouncing the first syllable of the word. Overall, Lea has to intervene three times before S7 is able to say, or repeat, the word “Pommes”. This example illustrates how eliciting specific information when a student is looking for a word can hinder progression in an interaction – indeed, it can evidently bring about the opposite of the intended effect, as seen in the setting above, when Lea’s intervention interrupts the interaction between the two students for a longer period of time and most likely also confuses S7.

During the test, Lea intervened to elicit complex grammar structures such as subordinate clauses with *damit* [so that] even more frequently than she did to cue individual words. These interventions always caused longer interruptions in a peer interaction because she had to intervene at least twice and generally had to correct student errors. Lea's primary goal was most likely to control whether the students were able to use a structure that had just been introduced during the unit. However, the intervention was not geared towards supporting peer interactions and therefore not in line with the main objective of paired speaking tests, namely assessing the interaction skills of students.

4.3 What benefits do teachers identify in summative assessments of speaking skills?

During the second research phase, the seven participating teachers were asked in retrospective interviews about how they would use the knowledge gained from the speaking tests ("Qu'allez vous faire avec les informations que vous avez obtenues grâce à cette épreuve?" [What will you do with the information you obtained through this test?]). When answering this question, most teachers began to speak about the feedback they would give their students after the test. Some teachers return the completed rubric including a mark (and sometimes a few comments on errors); however, all these teachers doubt whether their students would use this feedback to support further learning. For this reason, one teacher says she explains in class how the score or mark is calculated. Two other teachers who also doubt whether their students would take in the written feedback give each of their students the opportunity to listen to the audio recording of the exam and receive individualised oral feedback. Other teachers also forgo providing written feedback; rather, they offer to discuss the test (the mark or the individual performance) with their students. In addition to providing individual feedback, a few teachers also say they speak to the entire class about common errors. Some teachers report that this kind of group feedback is a way to encourage particularly anxious students and to boost self-confidence in general. Two teachers also say that they plan to use the information gained to revise the test task and/or the assessment rubric. Another teacher made a spontaneous decision during the exam to not rate this performance; rather, he would repeat the test and prepare his students better.

When asked whether they think speaking tests are useful for helping students learn German, the same seven teachers were divided, with roughly half saying yes and half no. Lea, for instance, sees no general benefit in oral exams: "c'est plus une source de stress ces évaluations que une source de progression" [these tests are more a source of stress than a way to learn]. She believes students can be better motivated to learn by arranging a meet-up with a partner class from German-speaking Switzerland. In the retrospective interview, one other teacher also expressed doubts as to whether her students would benefit from oral exams, as young people their age do not yet see the point of learning German. Two other teachers differentiate between the lower and higher tracks: in the case of lower-track classes – with which they participated in the study – they see fewer advantages of oral

exams, as these students rarely study for the tests; by contrast, students in the higher track generally prepare for exams. The remaining three teachers have a more positive attitude towards oral exams. Eva and Tom, whose lower-track classes participated in the study, see tests mainly as a way to motivate (or force) students into preparing and studying more seriously. The third teacher has a different line of argument: he believes oral exams are the tests with the greatest learning effect because they help students realise what their language level actually is and whether they are able to interact in everyday situations. This in turn either boosts the students' confidence or it shows them where they need to make a greater effort. At the same time, this teacher also likes oral exams because the setting enables him to offer his students more targeted support than in a written examination and allows him to give them oral feedback immediately following a test.

On the whole, a certain sense of resignation concerning the value of oral exams became apparent during the teacher interviews, in particular regarding whether their feedback can bring about a lasting learning effect, and a certain helplessness as to how they could make improvements was evident. Most teachers expressed doubts that students would incorporate and use written feedback (in the form of a filled in rubric, for example) for future learning. And although past research has shown that feedback is particularly conducive to learning when it is given soon after an exam, when it is concrete and descriptive, and when it helps learners see what and how they can improve (Luoma, 2004; Hattie & Timperley, 2007), this kind of concrete (oral) feedback given immediately after a test was observed only to a limited extent in the case of two teachers – presumably due to time constraints and other practical reasons. It is therefore unsurprising that roughly half of the seven teachers interviewed take a rather critical view of the benefits of speaking tests for various reasons, especially in classes in the lower track.

4.4 What needs do teachers have regarding the summative assessment of speaking skills?

Over the course of the teacher interviews, it became evident that conducting speaking tests gives rise to various difficulties. For instance, these types of tests are time-consuming, especially when teachers record the oral tests and assess student achievement only afterwards. In addition, the teachers frequently mentioned disciplinary issues that make it impossible for them to test students outside the classroom, as the rest of the class cannot be left unattended. With this in mind, the twelve teachers from the first phase of the study were asked if additional support or materials for the summative assessment of speaking skills would be desirable. The teachers from Fribourg and Neuchâtel gave very different answers to this question, which is likely due to the different framework conditions in these cantons. In the Canton of Fribourg, teachers are required to test the speaking skills of their students several times per class and year (depending on the number of weekly lessons); they must also use a predefined assessment rubric with four criteria: task, communication, formal accuracy and pronunciation/intonation. By contrast, the Canton of Neuchâtel merely

recommends that teachers test oral skills but provides no details on how frequently this should happen or what criteria should be used to make an assessment. When answering the question about any desired support, the four teachers from Neuchâtel said at various times that their university of teacher education provides useful material for assessments. Nevertheless, they would welcome specific, ready-to-use materials for assessing speaking skills as well as materials that could be adapted for different groups.

When the Fribourg teachers were asked about their needs for support in assessing speaking skills, all eight spontaneously offered their opinion on the compulsory rubric. Four were very critical of the rubric, two criticised certain aspects and the remaining two were satisfied with the assessment instrument. The critical teachers were unanimous in saying that the rubric was too general and that it should be made more specific. Some also pointed to difficulties in interpreting the categories: “erfüllt/nicht erfüllt beim Raster – was heisst das konkret?” [“target met/not met” in the rubric...what does that actually mean?] (Eva), while others say it was too subjective: “Des fois j’ai l’impression que c’est une appréciation très personnelle finalement» [at times I have the impression that, in the end, it very much depends on who is giving the marks] (teacher 06), or “On évalue théoriquement tous de la même façon. Voilà je dis bien théoriquement parce que je suis pas tout à fait convaincue de ça” [In theory, we’re all evaluating the same way. But I say ‘in theory’ because I’m not convinced everyone does it the same] (teacher 04). Some also voiced more fundamental criticism; for instance, Lea believes applying the rubric is unfair to her students, as it would assess them in a deficit-oriented way:

J’aime pas vraiment travailler avec ces grilles cantonales. Je trouve que je considère pas assez l’élève et que je me concentre sur ces quatre critères et que chaque fois qu’il y a un truc qui coince je me dis ah ben là j’enlève un point. Puis c’est vraiment j’ai l’impression une évaluation vers le bas : plutôt que de valoriser ce qu’ils ont déjà su [...]. J’ai pas l’impression que c’est le meilleur outil qu’on ait pour évaluer l’oralité.

[I don’t like working with the cantonal rubric. I find that I don’t consider the students enough and that I’m concentrating on these four criteria and that each time there is something that is not quite right, I tell myself I take off one point. I really get the impression that an evaluation tends to look at what’s wrong: rather than valuing what they know (...). I don’t have the impression that it’s the best tool we would have for assessing oral language skills.]

Most of the interviewed teachers in the Canton of Fribourg would like the rubric to be both more specific and more nuanced. They also say it would be helpful if it provided examples of assessment practices – or that using it was no longer mandatory. And if they could choose additional resources, they would prefer clearer and more detailed materials for *practising* speaking.

The criticism levelled at the Fribourg rubric appears to be justified, as the instrument is indeed in various respects too vague and has other shortcomings as well. For one, the criteria leave a great deal of room for interpretation, as the rubric makes no reference to the competences defined by the CEFR or the curriculum and because it provides no examples (i.e. recordings of students that are commented on and rated). For example, because the values for the criterion “pronunciation and intonation” (generally correct, sufficiently correct, only partially correct, rarely correct, not correct) have no examples or definitions, using them for an assessment will be arbitrary by nature. Unsurprisingly, three teachers spontaneously admitted that they give all their students the same (sufficient) number of points for “pronunciation and intonation”, as they find it difficult to assess this criterion. They say their difficulties here stem from the vague nature of the rubric, but also because their students simply do not speak enough to enable an adequate assessment of their performance. Another weakness in the rubric is the grouping of very different aspects under the criterion “communication” – including “varied sentence structure”; “varied vocabulary”; “realisation of speaking intention”; “correct phrasing”; and “fluency” – which are aspects that the Lingualevel rubric, for example, assigns to separate criteria: “spectrum”, “accuracy” and “fluency” (Lenz & Studer, 2007). In addition to taking issue with the rubric itself, various teachers said they had received little or no training in working with the rubric, further calling the instrument’s reliability into question.

4.5 What do students think about speaking tests?

In a written survey ($n=254$) and retrospective interviews ($n=84$), students were also asked to share their views and wishes concerning speaking tests. The key statements are presented below (for more information on the student standpoint, cf. Lüthi et al., submitted).

In the questionnaire, the students were asked to indicate how useful they think the various test formats are for improving their German skills (Figure 3). As seen in their responses, the students think different kinds of formats are useful, with the paired speaking tests being rated most useful (80% [somewhat] agree), ahead of individual exams and presenta-

tions. In the interviews, the students were also asked whether they think speaking tests help them make progress in German. On the whole, 57% ($n=48$) and 23% ($n=19$) of the students responded that tests are useful, or at least partially useful, while 18% ($n=15$) said they are not useful. Students with a positive opinion generally said that speaking tests help them develop the speaking skills (14 students, 17%) that they will need to speak German later in life (e.g. for an apprenticeship or work) or when travelling in German-speaking areas (10 students, 12%), or that the tests are a good way to practise pronunciation (9 students, 10%). In general, many of their arguments tend to say that speaking tests offer an opportunity to practise speaking or aspects of spoken language, and that speaking tests are useful because speaking is an important and useful competence. The students who responded that speaking tests have no value gave various reasons for this opinion, the most common of which was test anxiety (6 students, 7%).

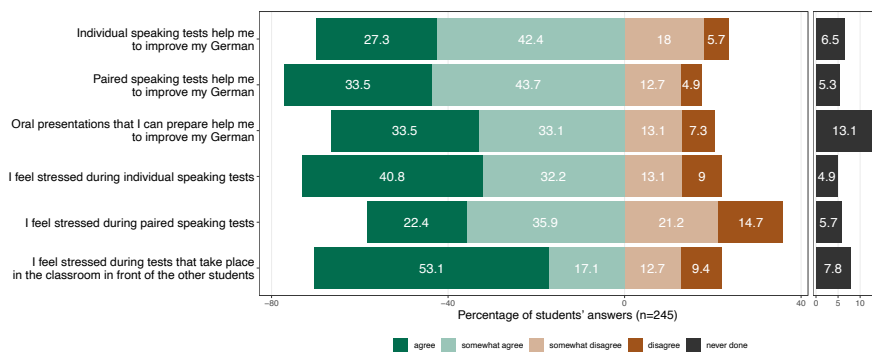


Figure 3: Results of the student questionnaire on perceived benefits and stress levels in different test formats

In the questionnaire, the students were also asked whether they experience test anxiety in the different kinds of oral test formats (Figure 3). It is striking to note that students experience exam anxiety even in test formats they find useful, and also that they rate paired speaking tests as the least stressful and most useful. Three quarters of the students rate individual exams as stressful or somewhat stressful; this percentage lowers to about 60% when a test is completed in pairs. When considering only “agree” answers (i.e. the number of stressed students) to the items “I feel stressed during individual/paired speaking tests” an even clearer picture emerges: whereas nearly 41% of the students report high stress levels during individual tests, the percentage is reduced by half (to 22%) for reported stress during paired speaking tests. In addition, many students report feeling very stressed when a test requires them to speak in front of the entire class, as indicated by the high percentage of agreement (53%) with the item “I feel stressed during tests that take place in the classroom in front of the other students”.

Students were also asked about their preferences with regard to feedback on speaking tests. Their responses showed an interest in all proposed feedback formats. Their preferences (ranked from highest to lowest) are: 1) suggestions for improving; 2) teacher remarks on what went well and less well on the test; 3) a mark; 4) a completed assessment rubric; 5) feedback on how the entire class performed on the test. These results reveal that student preferences for feedback and the teacher feedback they actually receive are only poorly aligned. Teachers always mark the tests, and they sometimes use an assessment rubric or give feedback to the entire class (cf. section 4.3). However, the feedback students are most interested in – suggestions for improving and individual remarks on what went well or less well – is either never or only very rarely given by the teachers surveyed.

In the questionnaire, the students also provided brief information about what they do with the feedback they receive. When analysing these responses, it is important to consider that not all students receive the same kind, or the same amount, of feedback. Nevertheless, their answers can be grouped in three categories.⁶ First, some students say they use teacher feedback to improve their German skills ($n=121$, 57%). Here, some of the student responses are very vague, for example: “je m’améliore” [I improve]; other students say they want to improve mistakes or weaknesses but give no indication of how they would go about doing this; still others refer to the effort they are making, for example: “Guardo dove devo migliorare di più e mi alleno a casa e provo a migliorare” [I look where I have to improve more and I practise at home and try to improve]. There are also students who say they will use the feedback (to prepare) for the next test, although here, too, most responses provide no details about how, or on what, they plan to work. Second, a number of students report that they look at the feedback ($n=66$, 30%), with some saying they focus on errors or weaknesses, while others provide no details on what they would look at. Third, there are students who say they do not use the feedback ($n=47$, 19%): some responded with “nothing”, others stated only that they put the feedback in a folder or show it to their parents. The responses suggest that many students understand that feedback can promote further learning. However, the often very vague statements also suggest that the students rarely (are able to) incorporate the feedback into their learning, thus also raising the question as to how actionable teacher feedback is for the students (cf. section 4.3).

6 Some answers were assigned both to the first category (use for future learning) and to the second (look at feedback).

5 Final remarks

The aim of the research project was to document formative and summative assessments of speaking skills in GFL lessons at the lower secondary level, and to collect data on teacher views on both assessment types as well as those of students on summative assessments. During the GFL lessons observed in the first project phase, spontaneous corrections were most frequently noted, while planned formative assessment activities were only rarely observed (cf. Davison & Leung, 2009, for an overview of formative assessment types). When assessment practices in the classroom are considered against the backdrop of their purpose to foster learning, it is furthermore striking that very little individual feedback that supports the development of speaking skills, or pointers on ways students could improve, were observed during the classroom observation sessions (or the summative tests), although it should also be stated that the study provided only a brief insight (two lessons) into an authentic classroom setting. Nevertheless, responses to the student questionnaire revealed that learners would indeed particularly appreciate suggestions on ways to improve their speaking skills (cf. section 4.5).

The second project phase focused on how speaking tests are conducted at the lower secondary level and how teachers deal with this complex task. When analysing the filmed, authentic speaking tests in eight classes, the high variability in teacher behaviours became apparent, particularly during the paired speaking tests. While one teacher (Tom) almost never intervened – even when students were having obvious interactional troubles – the test situation almost evolved into a German lesson in the case of another teacher (Lea), who intervened in the paired tests frequently and for different reasons: to manage the test when interactional troubles arose, to correct students, to provide positive feedback, but also to check whether students could actively use recently taught grammatical structures. Conversation analysis of these paired speaking tests revealed that not all interventions had a positive impact on the student interactions. Whereas prompts to improvise or reformulations of an unclear question promoted the progression of a peer interaction, eliciting words and structures in particular led to longer interruptions in the exchange. Hence, eliciting is contrary to the actual aim of paired speaking tests, which is testing a student's capacity to engage in an oral interaction. Moreover, the practice of eliciting also means that the advantages of paired speaking tests cited in specialist literature – that learners assume more responsibility and are less reactive than in individual tests, for example (Taylor & Wigglesworth, 2009; Rydell, 2019) – do not come to bear. Various explanations for the differing examiner behaviours observed can be considered, including different test purposes. For instance, Tom uses a proficiency test that was not part of the regular class textbook to test oral interaction skills, whereas the other three teachers are interested in determining whether their students have met the learning objectives of a textbook unit, including the ability to use specific vocabulary and grammatical structures. For this reason, the assessment of interaction skills was most likely not foregrounded in these achievement tests,

which is contradictory in itself, as the tests nevertheless remain speaking tests. As a rule, it was observed that teachers pursued different, even divergent, aims with the paired speaking tests: eliciting correct grammar structures, for example, has little to do with testing interaction skills.

Another core reason for the different examiner behaviours may be that teachers – unlike in standardised tests – must perform several roles during an exam: they are test administrators and raters but also, simultaneously, the students' teacher and at times their "interlocutors" (Teasdale & Leung, 2000; Koizumi, 2022). The analysis results suggest that the four teachers also weighted these roles differently: whereas Tom remained mainly in the roles of rater and test administrator, Lea often took on the role of a teacher who elicits and occasionally corrects grammar. Moreover, because Lea, Eva and Mia designed tests that do not meet the quality criteria of communicative tasks, they also had to assume a more active role in exam management (organising the candidates' turn-taking, for example). Hence, the study findings also indicate that teachers should be made aware of these various roles, and that they should understand they neither can nor should fill all of them equally well and at the same time. For instance, the interlocutor role (and not simultaneously that of rater) is important during an individual test, but exam management (and possibly rating) should be the focus during paired tests.

Although the assessment of student performance was not the focus of the study, the teacher survey provided clear evidence that teachers are certainly critical of the assessment criteria used, but – in the case of teachers in the of Canton of Fribourg – they are obliged to work with a specific assessment rubric. This rubric is a one-size-fits-all instrument used for both paired and individual speaking tests (e.g. presentations) and is certainly rightly criticised by the teachers for lacking clarity and nuance. In the same vein, criticisms that assessments made using the rubric are subjective also seem justified: because the rubric makes no reference to the CEFR language proficiency levels and provides no examples, assessments are inevitably based on a teacher's personal experience and subjective standards.

In the interest of ensuring the fairest possible assessment of student performance, a general need to better align learning objectives, test purpose, test tasks, teacher behaviours and assessment criteria was discerned in the tests observed in the study. At the same time, however, it should be noted that teachers are also bound by external conditions (such as limited time and staffing resources or, depending on the canton, directives and proscriptions) that cannot be said to simplify their complex work. To best support teachers in assessing speaking skills, it would be beneficial if – especially in the Canton of Fribourg – teachers were given greater freedom to independently decide how they control whether their students have met learning objectives. This includes choosing an assessment rubric that is adapted to the test format and purpose, but also the ability to test vocabulary and grammar from time to time. This more flexible approach means teachers would no longer need to intervene during an oral exam only for the sake of eliciting grammatical structures that are unrelated to, and unnecessary for, performing the communicative task. Al-

though the teachers themselves say they have enough material for tests, the study findings point to a need for teacher training and professional development in several areas, in particular in identifying different test purposes, in task design, but also in providing learning-oriented feedback and in intervention behaviours. To help teachers gain a better understanding of how their actions impact students and learning, the videos created and pre-processed for the project can be made available to other researchers on request.

Some of the filmed speaking tests were pre-processed for teacher training and professional development purposes.

Individuals working at higher education institutions are invited to contact Moritz Sommet (moritz.sommet@unifr.ch), head of the documentation centre at the Research Centre on Multilingualism (RCM), to request these videos. The accompanying document (which includes transcripts, information on test tasks and expert rating of student performance, in German) is available at <https://centre-plurilinguisme.ch/en/research/assessment-speaking>.

**6 Bibliografie
Bibliographie
Bibliografia
Bibliography**

- Berry, V., Sheehan, S. & Munro, S. (2019). What does language assessment literacy mean to teachers? *ELT Journal*, 73(2), 113-123. <https://doi.org/10.1093/elt/ccy055>
- Council of Europe. (2018). *Common European framework of reference for languages: Learning, teaching, assessment. Companion volume with new descriptors*. Council of Europe. <https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989>
- Csépes, I. & Fekete, A. (2018). Assessing speaking skills. In: Tsagari, D., Vogt, K., Froelich, V., Csépes, I., Fekete, A., Green, A., Hamp-Lyons, L., Sifakis, N. & Kordia, S. (eds.), *Handbook of Assessment for Language Teachers*. 98-128. <http://taleproject.eu/>
- Davison, C. & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly*, 43(3), 393-415. <https://doi.org/10.1002/j.1545-7249.2009.tb00242.x>
- De Jong, N. H. (2023). Assessing second language speaking proficiency. *Annual Review of Linguistics*, 9(1), 541-560. <https://doi.org/10.1146/annurev-linguistics-030521-052114>
- EDK. (Schweizerische Konferenz der kantonalen Erziehungsdirektoren) (2011). *Grundkompetenzen für die Fremdsprachen. Nationale Bildungsstandards. Frei gegeben von der EDK-Plenarversammlung am 16. Juni 2011*. Bern: EDK. https://edudoc.ch/record/96780/files/grundkomp_fremdsprachen_d.pdf
- Europarat. (2001). *Gemeinsamer europäischer Referenzrahmen für Sprachen: Lernen, lehren, beurteilen* (J. Quetz & G. Schneider, Übers.). Langenscheidt.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132. <https://doi.org/10.1080/15434303.2011.642041>
- Galaczi, E. D. (2014). Interactional competence across proficiency levels: How do learners manage interaction in paired speaking tests? *Applied Linguistics*, 35(5), 553-574. <https://doi.org/10.1093/applin/amt017>
- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research* 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385-402. <https://doi.org/10.1177/0265532208090158>
- Isaacs, T. (2016). Assessing speaking. In Tsagari, D. & Banerjee, J. (eds.), *Handbook of second language assessment*. Boston: De Gruyter, 131-146. <https://doi.org/10.1515/9781614513827>

- Kamiya, N. (2016). The relationship between stated beliefs and classroom practices of oral corrective feedback. *Innovation in Language Learning and Teaching*, 10, 206-219. <https://doi.org/10.1080/17501229.2014.939656>.
- Koizumi, R. (2022). L2 Speaking assessment in secondary school classrooms in Japan. *Language Assessment Quarterly*, 19(2), 142-161. <https://doi.org/10.1080/15434303.2021.2023542>
- Lenz, P. & Studer, T. (2007). *Lingualevel. Instrumente zur Evaluation von Fremdsprachenkompetenzen*. Bern: Schulverlag.
- Luoma, S. (2004). *Assessing speaking*. Cambridge University Press.
- Lüthi, G., Peyer, E. & Ravazzini, N. (eingereicht). Speaking tests in the lower secondary foreign language classroom: students' views on usefulness, anxiety, preparation and feedback.
- Lyster, R. & Ranta, L. (1997). Corrective feedback and learner uptake. Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, 19(1), 37-66. <https://doi.org/10.1017/S0272263197001034>
- Lyster, R. & Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition*, 32(2), 265-302. <https://doi.org/10.1017/S0272263109990520>
- May, L. (2009). Co-constructed interaction in a paired speaking test: The rater's perspective. *Language Testing*, 26(3), 397-421. <https://doi.org/10.1177/0265532209104668>
- Mayring, P. (2010). *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. (11., aktualisierte und überarbeitete Auflage). Weinheim: Beltz.
- Peyer, E., Lüthi, G. & Ravazzini, N. (2023). Formative Beurteilung des Sprechens im DaF-Unterricht: Praktiken und Überzeugungen von Lehrpersonen der Sekundarstufe I. *Informationen Deutsch als Fremdsprache*, 50(5), 492-504. <https://doi.org/10.1515/infodaf-2023-0082>
- Peyer, E., Lüthi, G. & Ravazzini, N. (eingereicht). Teacher reactions to interactional troubles during classroom-based paired speaking tests at lower level: a trade-off between progressivity and candidates' independent interactional trouble solving.
- Roothoof, H. (2014). The relationship between adult EFL teachers' oral feedback practices and their beliefs. *System*, 46, 65-79. <https://doi.org/10.1016/j.system.2014.07.012>

Rydell, M. (2019). Negotiating co-participation: Embodied word searching sequences in paired L2 speaking tests. *Journal of Pragmatics*, 149, 60-77. <https://doi.org/10.1016/j.pragma.2019.05.027>

Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing*, 30(3), 309-327. <https://doi.org/10.1177/0265532213480128>

Schegloff, E. A. (2007). *Sequence organization in interaction*. Cambridge University Press.

Selting, M. et al. (2009). Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion*, 10, 353-402. <http://www.gespraechsforschung-online.de/fileadmin/dateien/heft2009/px-gat2.pdf>

SeNOF (Service de l'enseignement obligatoire de langue française, état de Fribourg) (2018). *Instructions du SeNOF relatives à l'enseignement et à l'apprentissage du vocabulaire dans les langues étrangères à l'école obligatoire*.

Taylor, L. & Wigglesworth, G. (2009). Are two heads better than one? Pair work in L2 assessment contexts. *Language Testing*, 23(6), 325-339. <https://doi.org/10.1177/0265532209104665>

Teasdale, A. & Leung, C. (2000). Teacher assessment and psychometric theory: A case of paradigm crossing? *Language Testing*, 17(2), 163-184. <https://doi.org/10.1177/026553220001700204>

Tsagari, D. & Vogt, K. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. *Papers in Language Testing and Assessment*, 6(1), 41-63.

Vogt, K. & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11(4), 374-402. <https://doi.org/10.1080/15434303.2014.960046>

Vogt, K. (2018). *Teachers' assessment literacy enhancement. Needs analysis report*. http://taleproject.eu/pluginfile.php/2127/mod_page/content/2/TALE%20Needs%20analysis%20report.pdf

7 Transkriptionszeichen

Conventions de transcription

Convenzioni di trascrizione

Transcription conventions

	Pause von unter einer Sekunde: Pause de moins d'une seconde : Pausa di meno di un secondo: Pause of less than a second:
(-)	ca. 0.2-0.5 Sek.
(--)	ca. 0.5-0.8 Sek.
(--)	ca. 0.8-1.0 Sek.
(1.0), (2.8)	Pause mit Angabe der Länge in Sekunden und Zehntelsekunden Pause mesurée en secondes et 10 ^{es} de secondes Pausa con durata in secondi e decimi di second Pause in seconds and tenths of a second
	Dehnung: Allongement : Allungamento: Prolonged vowel or consonant:
:	ca. 0.2-0.5 Sek.
::	ca. 0.5-0.8 Sek.
((schaut zur LP)), ((nickt))	Nonverbale Handlungen Gestes ou actions non verbales Azioni non verbali Non-verbal actions
je [sais] [oder kennst] du	Überlappungen und Simultansprechen Chevauchement Parlato simultaneo o sovrapposto Overlapping speech
schu:\ sch\	Abbruch Troncation d'un mot (Auto)interruzione Cut-off
GENAU, KAUFen	Betonte Wörter oder Silben Accentuation d'un mot ou d'une syllabe Enfasi su una parola o una sillaba Emphasis on word or syllable

hm_hm	Rezeptionssignal (bejahend) Signal de réception (affirmatif) Segnale di ricezione (affermativo) Reception signal (affirmative)
<non>, <help>	Wörter in anderer Sprache als Deutsch Mots dans une autre langue que l'allemand Parole in una lingua diversa dal tedesco Words in a language other than German
(c'est)	Vermuteter Wortlaut Mot incertain Trascrizione incerta Dubious transcription
(unv.) (non-int.) (incompr.) (incomprehensible)	Unverständliche Passage, teilweise mit Angabe der Dauer Passage non-intelligible, parfois avec indication de la durée Passaggio incomprensibile, a volte con indicazione di durata Unintelligible speech, sometimes with indication of duration

