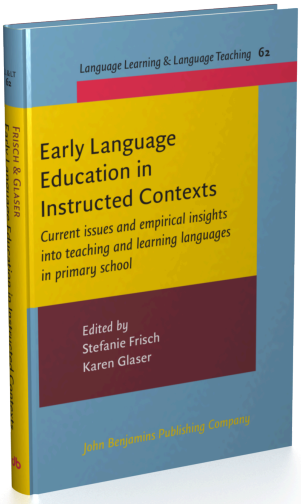
Language Learning & Language Teaching 62

CHAPTER 12

Student teachers' and mentors' perceptions of effective teaching techniques in the primary L2 English classroom

Kristin Kersten | Mannheim University
Karen Glaser | Leipzig University
Hannah Ruhm | University of Potsdam
Jana Roos | University of Potsdam
Sonja Brunsmeier | University of Passau
Martin J. Koch | University of Würzburg

https://doi.org/10.1075/lllt.62.12ker



Pages 260–285 of
Early Language Education in Instructed Contexts: Current issues and empirical insights into teaching and learning languages in primary school
Edited by Stefanie Frisch and Karen Glaser
[**Language Learning & Language Teaching**, 62]
2025. xiv, 330 pp.

John Benjamins Publishing Company

CHAPTER 12

Student teachers' and mentors' perceptions of effective teaching techniques in the primary L2 English classroom

Kristin Kersten[1], Karen Glaser[2], Hannah Ruhm[3], Jana Roos[3], Sonja Brunsmeier[4] & Martin J. Koch[5]
[1] Mannheim University | [2] Leipzig University | [3] University of Potsdam | [4] University of Passau | [5] University of Würzburg

This chapter presents the results of a multi-site study that employed the *Teacher Input Observation Scheme* (TIOS, Kersten et al., 2018) in the context of the primary second language (L2) English practicum to elicit self- and other-assessments of classroom practices from 74 student teachers and their 13 university mentors at three German universities. The four TIOS sub-scales permit a differentiated view on mentor and student perceptions of language teaching strategies to shed light on this crucial phase of pre-service teacher education programmes. ANOVAs showed that student and mentor ratings did not differ systematically, but a significant interaction was found between ratings and universities, hinting at a possible intervening role of academic programs. Linear regression analyses revealed that pedagogic experience negatively predicted ratings, suggesting that both groups' assessments become stricter with teaching experience. Implications for primary English language teacher education and research are discussed.

1. Introduction

Research in Instructed Second Language Acquisition (ISLA) has identified multiple L2 classroom strategies that have been shown to promote learners' L2 development. Among these, the quality of L2 input plays a central role (Gass et al., 2020), including cognitively stimulating activities (Leow, 2015; Praetorius et al., 2018, 2020), tasks (Ellis et al., 2020; Long, 2015) and opportunities for output production (e.g., Swain, 2005; for an overview, see Loewen & Sato, 2018). This is especially relevant in the primary L2 classroom, as language development in young learners relies heavily on implicit learning and thus exposure to rich and meaningful input (Muñoz, 2019). Accordingly, the planning and delivery of lessons

https://doi.org/10.1075/lllt.62.12ker

that feature high-quality input are skills targeted in teacher education, most centrally during the teaching practicum, in which student teachers receive feedback from their university mentors. Yet, little research has been carried out to explore whether students' perceptions of the instructional quality of their own lessons correspond to those of their mentors. The present study addresses this gap and investigates how teaching techniques in the primary English teaching practicum are evaluated by pre-service teachers ('students') and their practicum mentors from university ('mentors'). Using the well-established Teacher Input Observation Scheme (TIOS, Kersten et al., 2018) to operationalise input quality, the study contrasts the self- vs. other-perceptions of 74 student teachers and 13 university practicum mentors at three German universities, focusing on their (diverging) perceptions of input quality, differences between the three universities, and the predicting effects of students' and mentors' previous pedagogical experience. In contrast to previous research in the field, which has mainly focused on general pedagogical aspects, our study thus addresses subject-specific aspects of L2 English teacher education. Identifying gaps between self-assessments and external evaluations in this context has the potential to lead to a stronger focus on specific concepts in academic courses and practicum preparation to improve mentoring quality and further primary language teacher education.

2. Theoretical background

2.1 Role and effects of input quality in ISLA

The present study investigated self- and other-perceptions of input in the primary English classroom, as the significance of input is considered a crucial factor and a fundamental driving force for Second Language Acquisition (SLA) (Gass et al., 2020). To that end, we adopt the wide definition of this concept suggested by Truscott and Sharwood Smith (2019), which includes all sensory stimulation in the learners' environment, encompassing visual, auditory, somatosensory, and other sensory information, i.e., "everything that contributes to the interpretation of an utterance and which can lead to further development of an individual's linguistic ability" (p. 10). In this definition, the authors include both the situational context — all external stimuli in the immediate environment that learners encounter during a situation — and the discourse context — stimuli from the actual linguistic interaction. Therefore, a comprehensive understanding of input needs to comprise at least the following: (1) the characteristics of the activities selected by the teachers and carried out by the learners (situational context), (2) teachers' verbal behaviour, (3) their non-verbal stimuli, and (4) the opportunities for learner output in these

activities and teachers' feedback to them during classroom interaction (discourse context). Teaching strategies pertaining to these four categories of input quality represent essential pedagogical skills that student teachers need to acquire. They are therefore comprised in the four scales of the TIOS.

There are several models in SLA research which support the effectiveness of L2 input quality in this broad sense as a predictor for L2 acquisition and, thus, their relevance for teacher education (e.g., Gass et al., 2020; Kormos, 2011; Leow, 2015). These models focus on processes of intake and storage of new (linguistic) information, in accordance with the following reasoning: All external sensory stimuli trigger the learner's processing systems. The actual 'intake', however, i.e., the internal construction of new knowledge, its storage, and subsequent output production depend on several characteristics of these input signals. Among these characteristics are the strength and frequency of the signals and, thus, their salience, which leads to noticing and awareness. They also activate internal networks of stored associations and schemas (prior knowledge) as well as emotional responses, which are, in turn, related to learners' interests and motivation. All these aspects are directly related to the depth of cognitive involvement (*deep processing*, Leow, 2018), the transformation of input into intake and storage (Li & Jeong, 2020), and helping learners monitor their own language use and state of interlanguage. (For more in-depth overviews of these processes, see Kersten, 2021, and Kersten et al., 2024). These factors were considered in the construction of the four TIOS scales (Kersten et al., 2018), which were employed in the present study. More specifically, they include characteristics of interactional classroom activities in which input is embedded (situational context) (scale 1), teachers' verbal input modifications (scale 2), their non-verbal input to scaffold comprehension (scale 3), and characteristics of output support (discourse context) (scale 4).

Following these models, the quality of L2 input encompasses various scaffolding strategies to increase comprehensibility, such as linguistic and interactional modifications to enhance cognitive stimulation and metalinguistic awareness (Kersten, 2021). Numerous ISLA studies show positive effects of L2 input quality on learners' L2 attainment and thus underscore its relevance for pedagogical practices (see the overview in Kersten, 2019; cf. also Ellis & Shintani, 2014; Loewen, 2020; Loewen & Sato, 2018; Long, 2015). Predictive effects were found, e.g., for L2 English in the DESI study (Helmke et al., 2008) with 105 9th-grade classes. This study was based on classroom videography and adapted several standardised instruments including the Communicative Orientation of Language Teaching Observation Scheme (COLT, Spada & Fröhlich, 1995; Spada, 2024; cf. also Spada, 2019, for a discussion of observation schemes in classroom research), which also served as background for the TIOS. The TEPS study *(Teaching English in Primary*

Schools) with 269 (Wilden & Porsch, 2019) and 766 learners (Wilden et al., 2020) reported comparable results.

Similarly positive effects of input quality on L2 development have been found in research using the TIOS and its precursor for preschools, the Input Quality Observation Scheme (IQOS, Weitz et al., 2010), which both operationalise L2 input quality by means of specific scaffolding techniques. Predictive effects are shown by Weitz (2015) with the IQOS for 210 young learners' L2 receptive skills in bilingual preschools, and for the TIOS by Kersten (2023) with 93 primary L2 learners, Kersten et al. (2021a) with 79 primary L2 learners, Kersten et al. (2023) with 690 (L2) and 836 (L1) primary and early secondary learners, and by Kersten et al. (2024) with 933 primary L2 learners, respectively. Moreover, Kersten et. al (2023) found for both L2 English and L1 German instruction that teaching techniques with high TIOS-ratings moderated the effect of socioeconomic status on learners' L2 and L1 skills (cf. also Nilsen et al., 2020). As these studies confirm the general effectiveness of input quality as operationalised in the TIOS, the present study adopted this assessment instrument to collect student and mentor perceptions in the teaching practicum to shed light on this crucial phase of primary English language teacher education and add to the under-researched area of student and mentor perceptions of instructional quality.

2.2 Student teachers' and experienced teachers' perceptions of instructional quality and teaching competences

Differences between student and more experienced teachers have been investigated mostly in the context of general teacher education (Bransford et al., 2000; Hogan et al., 2003). For instance, Tsui (2009) identified differences between novice and experienced teachers in the ability to integrate knowledge, to respond to teaching contexts, and to engage in reflection. Other studies found differences in novice and experienced teachers' professional knowledge, e.g., when planning and conducting lessons, assessing learning difficulties, in their reflective practices (Schempp et al., 1998), or when processing and responding to classroom management situations (Wolff et al., 2016; 2021).

In contrast, comparisons of student teachers' perceptions of their own teaching with the evaluation by mentors are scarce, as research to date has mainly contrasted students' self-assessments before and after a teaching practicum. Although focusing on varying aspects of (perceived) teaching competence, they consistently found that students exhibited positive self-assessments, which increased during the practicum. Schubarth et al. (2012) investigated the self-perceptions of 144 student teachers in teaching, educating, assessing, advising, innovating, professional and methodological expertise, and found positive self-perceptions in all

these areas. Gröschner et al. (2013) assessed the development of self-perceptions in teaching, educating, assessing and innovating of 221 students during their practicum semester, and found a positive development in all four areas. Likewise, Festner et al. (2018) collected self-assessments of 810 student teachers at three different universities and found positive self-evaluations in instructional competence and teaching quality. This is in line with studies on teacher self-efficacy showing that student teachers' efficacy beliefs tend to increase during practical phases at school (for a summary cf. Schüle et al., 2017).

The few studies which focused on self-and other-perceptions of teaching competence yielded varied findings. On the one hand, some studies have found correlations between students' self-assessments and mentors' assessments, usually in the form of the mentors confirming the students' positive self-perceptions. For instance, Bach (2013) investigated the development of lesson planning competence in 488 students during their teaching practicum, showing that the students' self-assessments of their planning competence increased significantly over time, with no significant differences from their mentors' evaluations. Similarly, Brodhäcker (2014) found no significant differences in 512 teacher trainees' self-assessments of various subjects and school types and their mentors' assessments in teaching practicums.

In contrast, other studies report on diverging perspectives between trainee and expert teachers, often in the form of students' "overestimating" their skills (Rothland, 2018, p. 482), i.e., rating their teaching competences significantly higher than their mentors. Schubarth et al. (2012) found that the assessments of 144 student teachers of various subjects and their mentors differed in teaching, educating, methodological skills and social competence, with the student assessments being more positive than their mentors'. Bogard et al. (2017) observed overconfidence in student teachers, with the teacher trainees rating their abilities in specific areas of literacy instruction significantly higher than their mentors. In Dassa and Nichols' (2019) study, undergraduate primary education student teachers rated their subject-related content knowledge and teaching ability consistently higher than their supervisors.

To explain such discrepancies, Rothland (2018, p. 482) refers to results of empirical studies on shifts in competence self-assessment, which show that "student teachers are already convinced to possess all necessary skills in the central fields of the teaching profession before they have even started their internship." In psychology, this phenomenon is known as the Dunning-Kruger effect, a bias effect which links metacognition and knowledge, stating that a person with low ability, expertise, or experience in an area tends to overestimate their skills in that field (Dunning, 2011). Schempp et al. (1998, p. 9) observed this for the teaching profession in their interview study with five novice and five experienced teachers:

"Interestingly, the novice teachers believed they knew most, if not all, they needed to know to teach well, while competent teachers believed they had much to learn about their craft."

In contrast, Bach (2013) and Rothland (2018) found that student teachers evaluated themselves more negatively than their mentors, thus underestimating their teaching competences (Bodensohn & Schneider, 2008; Moser & Hascher, 2000). In a rare longitudinal study with 401 primary education students, Bodensohn and Schneider (2008) compared students' self-assessments and their mentors' assessments of pedagogical content knowledge and social and methodological competences, finding that mentors rated students' competences more positively, while students were more self-critical in their ratings. Thus, in some contexts, student teachers receive rather positive or benevolent feedback from their mentors (Rothland, 2018). As Bach (2013) suggests, this could result from a close personal relationship developed during the practicum. Accordingly, the intensity of collaboration might be a mediating factor in the self-and other-perceptions of students' teaching skills.

As this review shows, the existing research into self- vs. other assessments of teaching competences yields contradictory findings. While some studies found significantly different ratings by experts and novices, in others these differed only in certain aspects or not at all. In addition, expert assessments have been found to be both less and more benevolent than the students' self-assessments. A possible reason for these discrepancies could be that the studies focused on different aspects (knowledge, teaching and/or reflection skills etc.), and mostly examined general rather than subject-specific areas of teaching competence.

Regarding the impact of teaching experience, numerous studies showed that students' self-assessed professional competences increase during school practicums, suggesting a positive correlation with practical teaching experience (Festner et al., 2018; Gröschner et al., 2013; Klingebiel et al., 2020; Schubarth et al., 2012). Some studies revealed that more advanced students with more general pedagogical knowledge and pedagogical content knowledge (König et al., 2018) show more positive self-ratings for teaching competence. In a rare study with prospective English teachers, Pfingsthorn and Gehring (2015) investigated changes in perceived teaching competences among 40 student teachers at different stages of their studies (BA, M.Ed.). Participants rated aspects of teaching competences in over 200 can-do statements in lesson planning, delivery, diagnostic assessment, and corrective feedback. Results showed that M.Ed. students rated themselves higher in areas related to planning and conducting lessons, but not in dealing with mixed-ability learners.

Overall, research into individual factors other than teaching experience that influence perceptions of teaching competences is scarce. This is especially true for

language education, resulting in a lack of research into perceptions of *language* teaching strategies and the matters of L2 input discussed in Section 2.1. What is more, the primary context is a particularly underexplored area in this regard. Addressing this context is all the more urgent as young learners' L2 development is especially dependent on high-quality input provision due to the fact that they learn more implicitly and exemplar-based than older learners (Muñoz, 2019). This illustrates the necessity of research into student teachers' self-perceptions and mentors' perceptions of L2 teaching quality in young language learner classrooms, a research gap which our study addresses.

3. Research aims and methodology

3.1 Research questions

To investigate students' and mentors' perceptions of L2 teaching quality in the primary English classroom, we carried out a cross-sectional multi-site study at three universities with 74 students and their 13 university mentors during the English teaching practicum, aiming at a detailed comparison of their differential perceptions of teaching techniques. Benefiting from the multi-site design, we also investigated possible differences in the ratings collected at the three teacher-education institutions. Students and mentors rated the student-taught lessons simultaneously using the TIOS prior to the lessons' reflection sessions. Specifically, the study addressed the following research questions:

RQ1: To what extent do student teachers and their university mentors differ in their TIOS ratings of students' teaching strategies in their primary English practicum lessons?
RQ2: To what extent do the TIOS ratings differ between universities?
RQ3: To what extent does pedagogic experience predict student and mentor ratings?
RQ4: To what extent do pedagogic experience and/or universities predict the differences between student and mentor ratings?

3.2 Sample

The sample consisted of 74 students from three German universities, Uni_1 ($n=21$), Uni_2 ($n=28$), and Uni_3 ($n=25$) and their 13 university mentors ('mentors', Uni_1: $n=3$, Uni_2: $n=4$, Uni_3: $n=6$). Mentors were academic staff employed at the three universities in the field of L2 English language teacher education and methodology. They were selected based on their institutional role as

university lecturers and EFL practicum mentors, and their training in and experience with mentoring student groups during their English teaching practicum (Table 1).

Table 1. Descriptive statistics of the mentor sample

		Uni_1 ($n=3$)		Uni_2 ($n=4$)		Uni_3 ($n=6$)		total ($n=13$)	
		n	%	*n*	%	*n*	%	*n*	%
Gender	male	0	0	1	25	1	17	2	15
	female	3	100	2	50	5	83	10	77
	non-binary	0	0	1	25	0	0	1	8
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age in years		51.33	11.15	37.75	6.65	39.00	3.74	41.46	8.34
University experience in years		11.00	2.00	7.50	2.65	5.42	3.68	7.35	3.64
School experience in years		14.67	5.69	0.38	0.48	5.25	2.48	5.92	6.13

All students were training to become primary school teachers (M.Ed. programmes at Uni_1 and 3, State Exam at Uni_2), and studied English as their first subject alongside others, mainly German, Math, and Science *(Sachunterricht)*. Most students identified as female, had no migration background, and only spoke German as their L1 (Table 2). Informed consent was obtained by all participants prior to the study.

Table 2. Descriptive statistics of the student sample

		Uni_1		Uni_2		Uni_3		total	
		n	%	*n*	%	*n*	%	*n*	%
Gender	male	2	9.52	3	10.71	2	8	7	9.46
	female	15	71.43	25	89.29	23	92	63	85.14
	missing	4	19.05	0	0.00	0	0.00	4	5.41
Migration Background	no	13	61.90	26	92.86	23	92.00	62	83.78
	yes	4	19.05	0	0.00	1	4.00	5	6.76
	missing	4	19.05	2	7.14	1	4.00	7	9.46
L1	German	10	47.62	26	92.86	24	96.00	60	81.08
	other L1	2	9.52	0	0.00	0	0.00	2	2.70
	German + other L1	3	14.29	0	0.00	0	0.00	3	4.05
	missing	6	28.57	2	7.14	1	4.00	9	12.16

Table 2. *(continued)*

		Uni_1		Uni_2		Uni_3		total	
		n	%	*n*	%	*n*	%	*n*	%
Subject	Math	8	38.10	0	0.00	3	12.00	11	14.86
	German	8	38.10	0	0.00	13	52.00	21	28.38
	Science	0	0.00	0	0.00	5	20.00	5	6.76
	Math, German, Science	0	0.00	28	100.00	0	0.00	28	37.84
	Physical education	1	4.76	0	0.00	2	8.00	3	4.05
	Music	0	0.00	0	0.00	1	4.00	1	1.35
	missing	4	19.05	0	0.00	1	4.00	5	6.76
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age		26.47	6.87	24.46	6.4	25.08	3.32	25.19	5.61
Teaching experience		4.33	1.53	6.88	12.62	8.71	12.01	7.57	11.91
Semester in the Master		2.18	0.39	0.32	0.67	3.08	0.88	1.74	1.41

Note. The number of Master semesters for Uni_2 was calculated starting with the 7th term of the State Exam programme. Teaching experience was operationalised as the months of substitute teaching at school.

3.3 Procedure

3.3.1 *Instruments*

Data were collected by means of the TIOS, a rating tool which operationalises input quality as several specific teaching techniques pertaining to L2 input, interaction learner activation, and output production. Derived from the cognitive-interactionist framework (Ellis & Shintani, 2014; Loewen, 2020; Loewen & Sato, 2018; Long, 2015) and pre-existing rating scales such as the COLT and IQOS (overviews in Kersten, 2021; Kersten et al., 2024), it comprises 41 *teaching techniques* defined as a "description of how a communicative behaviour or activity is carried out in the classroom at a given moment as the actual point of contact with the learner/s" (Kersten, 2021, p. 42). These are subsumed under the four following scales, which are illustrated by a few example items:[1]

1. For the full version of the TIOS, see https://www.researchgate.net/publication/340096869_Teacher_Input_Observation_Scheme_TIOS_and_Manual.

1. *Cognitively stimulating tasks/activities*
   Item 8: Tasks/activities are based on the prior world knowledge of the learners (i.e., their everyday experiences).
   Item 9: Tasks/activities include all learners actively at all times.

2. *Verbal input*
   Item 19: The teacher uses repetitions of key words and phrases.
   Item 23: The teacher uses intonation to stress key words/phrases in the L2.

3. *Non-verbal input*
   Item 26: The teacher uses body language.
   Item 30: The teacher provides displays (words/phrases/materials) in the L2 within the classroom.

4. *Support of output*
   Item 31: The teacher waits for learners' reactions/answers.
   Item 40: The teacher prompts learners' self-correction during interaction (i.e., encourages them to correct themselves).

All items are rated on a 0–5 Likert-Scale from 'not present at all' to 'present to a high degree'. The TIOS was extensively pilot-tested over six years to ensure content and construct validity, reliability and feasibility of the ratings (for a detailed description, see Kersten, 2020, 2021, 2023; Kersten et al., 2024).

Scale 1 (Tasks) comprises thirteen techniques and focuses on the characteristics of classroom activities. These determine the authenticity, meaningfulness and salience of the linguistic content, learners' meaningful interactions, problem-solving, prior knowledge requirements, active engagement, noticing of linguistic aspects, opportunities for interaction and production. The twelve items of Scale 2 (Verbal Input, abbreviated as VInput) focus on the adaptation of the teacher's language input to the comprehension level of the learners. Input modifications include the frequency of linguistic structures, lexical and structural variety, clear articulation, intonation, adapted speech rate, and teachers' level of L2 proficiency. Scale 3 (Nonverbal Input / NVInput) has five items and refers to non-linguistic techniques that are aimed at enhancing comprehensibility, i.e., body language, visual illustrations, manipulatives, classroom displays, and written labels. Scale 4 (Output Support / Output) comprises eleven items and focuses on teachers' reactions to learners' utterances through various forms of feedback, including praise, explicit corrective feedback, recasts, self-correction prompts, and waiting-time. For this study, the German self-report version of the TIOS was used (Kersten et al., 2021b).

An additional questionnaire collected background data on students' number of Master terms in their study programme, their months of substitute teaching experience, and mentors' number of years teaching at school and at university.

3.3.2 *Elicitation*

Data were collected between October 2021 and January 2023 during the practicum semester (Uni_1 and Uni_3), in which students spend 3–4 days per week at school, and the Weekly Practicum (Uni_2), in which students teach at a school for one day per week.

The TIOS was introduced shortly before the start of the practicum. Students read the TIOS and the accompanying manual, and rated two transcribed excerpts of video-taped lessons, which were subsequently discussed in class. Additionally, the theoretical background of the TIOS was introduced with identical training materials in all universities. The same material was given to the university mentors who observed the students at school.

Data elicitation took place during the classroom visits, where the TIOS was filled out by both students and mentors immediately after the lesson (Uni_1 and Uni_3), or after two consecutive lessons taught by two different students (Uni_2). The subsequent reflection and feedback sessions were also based on these observations, but no ratings were changed.

4. Results

4.1 ANOVA results for RQ1 (differences students-mentors) and RQ2 (differences between universities)

To address RQ1 (differences between students and mentors) and RQ2 (differences between universities), we calculated a repeated-measures analysis of variance (ANOVA, Table 4 and Figure 1) using IBM SPSS 28. Students and mentors were entered as a repeated measures factor, the universities as an additional between-subject factor. The ANOVA tests for mean differences between student teachers and EFL mentors averaged across all universities (RQ1), differences between all universities averaged across student teachers and mentors (RQ2), and the interaction of both grouping variables (students and mentors, universities). This means that it can identify whether student teachers and mentors only differ in one university (RQ1) or if universities differ regarding student teachers or mentors, or both (RQ2). Group means are displayed in Table 3.

We found no significant difference between students and mentors (all $p > .05$), neither for any of the scales nor for the TIOS Total Score, but significant group

Table 3. Group means (*M*) and standard deviations (*SD*) for all universities and student and mentor ratings for all TIOS subscales and the TIOS Total Score

	Uni 1 ($n=21$)				Uni 2 ($n=28$)				Uni 3 ($n=25$)			
	Students		Mentors		Students		Mentors		Students		Mentors	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Tasks	71.49	15.10	62.99	19.85	66.27	9.32	57.91	16.41	60.23	10.03	65.34	16.38
VInput	70.45	14.44	61.94	17.15	73.16	10.90	77.05	14.52	61.29	10.05	70.27	18.26
NVInput	68.67	20.51	65.00	21.08	74.95	14.29	70.86	14.10	67.08	14.91	70.88	16.50
Output	68.30	15.81	62.08	16.01	71.38	10.28	70.28	12.63	61.01	13.12	66.77	15.91
Total Score	**70.07**	**14.06**	**62.48**	**17.00**	**70.64**	**8.13**	**68.40**	**10.85**	**60.81**	**8.32**	**67.78**	**12.61**

Table 4. Results of the two-factorial ANOVAs for all TIOS subscales and the TIOS Total Score comparing students vs. mentors and all universities

	Students vs. mentors			University			Interaction		
	F(1, 71)	*p*	η^2_{part}	*F(2, 71)*	*p*	η^2_{part}	*F(2, 71)*	*p*	η^2_{part}
Task	3.13	.081	.04	1.39	.256	.04	4.24	.018*	.11
VInput	0.38	.540	.01	6.89	.002**	.16	4.49	.015*	.11
NVInput	0.24	.624	.00	1.56	.217	.04	0.94	.395	.03
Output	0.06	.812	.00	3.32	.042*	.09	2.46	.093	.07
Total Score	0.24	.625	.00	2.54	.086	.07	4.57	.014*	.11

Note.
* $p>.05$, ** $p>.01$, *** $p>.001$

differences between the universities for Verbal Input and Output Support. Significant interactions were identified for Tasks, Verbal Input, and the Total Score, indicating simple main effects for one of the grouping variables. Output Support showed a significant main effect for university without a significant interaction. We thus had a look at Bonferroni-corrected post-hoc tests to identify at which university the mean scores differed significantly. We found that the values at Uni_2 were significantly lower than those at Uni_3 ($p=.047$). No other significant differences were identified.

To further investigate the differences between students and mentors (RQ1) for Tasks, Verbal Input, Output Support and the Total Score for each university separately, we calculated conditional effects (simple main effects) using *t*-tests. Results showed that in Uni_2, student ratings for Tasks were significantly higher than the mentors' ($t(27)=2.38$, $p=.025$), and in Uni_1 this difference was nearly significant ($t(20)=1.74$, $p=.097$). In Uni_3, student ratings for Verbal Input ($t(24)=-2.19$,

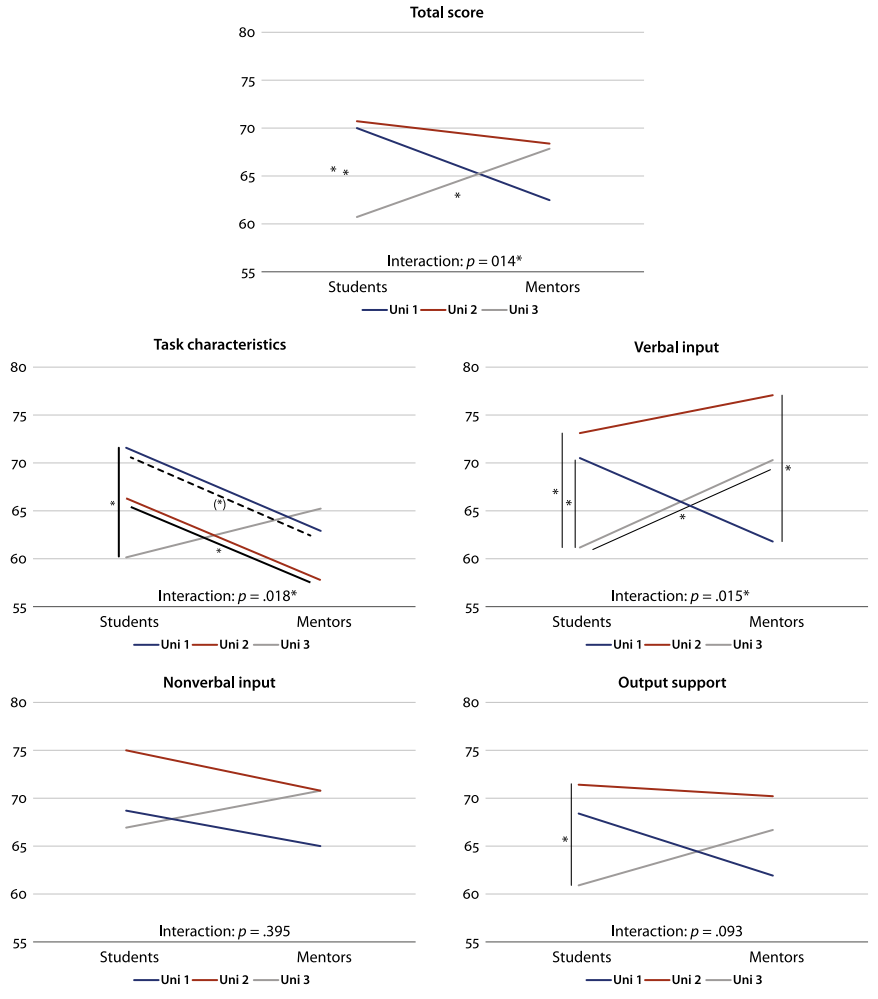


Figure 1. Group means for all universities and student and mentor ratings for all TIOS subscales and the TIOS Total Score

$p = .039$) and the Total Score ($t(24) = -2.32$, $p = .029$) were significantly lower than the mentors' ratings.

We then tested for differences between universities for students and mentors separately for each of the variables for Task, Verbal Input, Output Support, and Total Score to find simple main effects. For students, we found significant differences for Tasks ($F(2, 71) = 5.57$, $p = .006$) with significantly higher scores for Uni_1 compared to Uni_3 ($p = .004$) in the post-hoc test, for Verbal Input ($F(2, 71) = 7.20$, $p = .001$) significantly higher scores for Uni_2 compared to Uni_3 ($p = .001$) and Uni_1 compared to Uni_3 ($p = .031$), but no differences between Uni_2 and Uni_1 ($p > .05$). The same was found for the Total Scores ($F(2, 71) = 7.32$, $p = .001$) with

significantly higher scores for Uni_2 compared to Uni_3 ($p = .002$) and Uni_1 compared to Uni_3 ($p = .009$), but no differences between Uni_2 and Uni_1 ($p > .05$). For mentors, the only significant difference we found was for Verbal Input ($F(2, 71) = 4.97$, $p = .010$) with a significantly higher score for Uni_2 compared to Uni_1 ($p = .007$).

4.2 Regression analyses for RQ 3 (role of pedagogic experience) and RQ 4 (role of pedagogic experience and university combined)

For RQ3, which inquired about the extent to which pedagogic experience predicts student and mentor ratings, we used linear regressions to predict students' and mentors' TIOS scores using their pedagogic experience. For students, this was operationalised as months of substitute teaching experience and the number of Master semesters in the study programme (see *Note,* Table 3, for Uni_2). We found significant predictive effects of the number of semesters on Verbal Input, Output Support and the Total Score, explaining 16.9%, 11.8%, and 18.2% of the variance, respectively. The negative beta-values indicate that a higher number of semesters is associated with lower ratings for Verbal Input, Output Support, and of the TIOS Total Score (Table 5).

Table 5. Results of five separate linear regressions for students on subscales and the Total Score using months of teaching experience and the number of Master semesters as predictors

	R^2	$F(2, 52)$	p		b	SE	β	t	p
Tasks	.068	1.83	.171	teaching experience	−0.02	0.12	−0.02	−0.17	.869
				number of semesters	−1.72	0.91	−0.26	−1.90	.064
VInput	.169	5.09	.010*	teaching experience	−0.18	0.12	−0.19	−1.44	.156
				number of semesters	−2.62	0.94	−0.36	−2.78	.008**
NVInput	.076	2.07	.137	teaching experience	−0.26	0.18	−0.20	−1.45	.155
				number of semesters	−1.87	1.37	−0.19	−1.36	.179
Output	.118	3.34	.044*	teaching experience	0.11	0.14	0.10	0.74	.462
				number of semesters	−2.75	1.10	−0.33	−2.50	.016*
Total	**.182**	**5.57**	**.007****	**teaching experience**	**−0.05**	**0.10**	**−0.07**	**−0.52**	**.608**
				number of semesters	−2.52	0.77	−0.42	−3.27	.002**

Note.
* $p > .05$, ** $p > .01$, *** $p > .001$.
The number of Master semesters for Uni_2 was calculated starting with the 7th term of the State Exam programme. Teaching experience was operationalised as the months of substitute teaching at school.

Table 6. Results of five separate linear regressions for mentors on subscales and the Total Score using years of teaching experience at university and at school as predictors

	R^2	$F(2, 70)$	p		b	SE	β	t	p
Tasks	.077	2.92	.061	university experience	0.61	0.30	0.25	2.07	.043
				school experience	−1.31	0.71	−0.22	−1.84	.070
VInput	.107	4.18	.019*	university experience	−0.72	0.30	−0.29	−2.43	.018*
				school experience	−0.51	0.71	−0.09	−0.73	.471
NVInput	.034	1.22	.302	university experience	−0.32	0.30	−0.13	−1.05	.298
				school experience	−0.55	0.72	−0.09	−0.77	.447
Output	.109	4.3	.017*	university experience	−0.19	0.25	−0.09	−0.75	.453
				school experience	−1.48	0.61	−0.29	−2.45	.017*
Total	**.064**	**2.4**	**.099**	**university experience**	**−0.11**	**0.23**	**−0.06**	**−0.46**	**.645**
				school experience	−1.04	0.55	−0.23	−1.88	.064

Note.
* $p > .05$, ** $p > .01$, *** $p > .001$

For the mentors, their years of teaching experience at school and university were used to predict their TIOS ratings (Table 6). Results showed that Verbal Input was negatively predicted by university experience, while Output Support was negatively predicted by the school experience, explaining 10.7% and 10.9% of variance, respectively. Longer teaching experience is thus associated with lower ratings for Verbal Input and Output Support.

To address RQ4, which inquired about the extent to which pedagogic experience and/or universities predict the differences between student and mentor ratings, we calculated the difference value between student and mentor ratings for each TIOS subscale and the Total Score. We then used University and the predictors from the previous linear regressions (students' number of semesters and months of substitute teaching experience, mentors' teaching experience at university and school) to predict these difference-values in separate linear regressions (Table 7). The University variable, a categorical variable was changed into dichotomous dummy variables to be included in the linear regression. Since in dummy coding, two universities can be compared to the third university but not to each other, we used Uni_3 as a reference category because all significant differences between universities (except for Verbal Input in mentors) were found between Uni_3 and one of the other two universities. We then recalculated the regressions using Uni_1 as a reference category, to identify any additional differences between Uni_1 and Uni_2. We only found an additional significant difference between Uni_1 and 2 for Tasks ($b = -68.75$, $SE = 33.73$, $\beta = -1.88$, $t = -2.04$, $p = .047$).

Table 7. Results of five separate linear regressions on the differences between students and mentors for all subscales and the Total Score using University and students' and mentors' experience variables as predictors

	R^2	$F(6, 46)$	p	Predictors	b	SE	β	t	p
Tasks	.272	2,86	0,019*	Uni_1 vs. 3	71,26	26,46	0,90	2,69	0,010**
				Uni_2 vs. 3	2,51	12,50	0,07	0,20	0,842
				number of semesters	1,66	3,08	0,14	0,54	0,592
				school exp. (students)	−0,05	0,20	−0,03	−0,24	0,813
				school exp. (mentors)	−3,94	1,73	−0,98	−2,27	0,028*
				university exp. (mentors)	−0,77	1,03	−0,11	−0,75	0,459
VInput	.239	2,4	0,042*	Uni_1 vs. 3	15,20	27,05	0,19	0,56	0,577
				Uni_2 vs. 3	36,12	12,78	0,99	2,83	0,007**
				number of semesters	8,25	3,15	0,69	2,62	0,012*
				school exp. (students)	−0,05	0,20	−0,03	−0,25	0,804
				school exp. (mentors)	1,48	1,77	0,37	0,84	0,406
				university exp. (mentors)	−1,28	1,05	−0,18	−1,21	0,232
NVInput	.080	0.67	0,676	Uni_1 vs. 3	−3,92	33,65	−0,04	−0,12	0,908
				Uni_2 vs. 3	27,98	15,89	0,68	1,76	0,085
				number of semesters	4,86	3,91	0,36	1,24	0,220
				school exp. (students)	0,00	0,25	0,00	0,01	0,996
				school exp. (mentors)	1,14	2,20	0,25	0,52	0,607
				university exp. (mentors)	−0,50	1,31	−0,06	−0,38	0,704
Output	.186	1,75	.130	Uni_1 vs. 3	−2,51	27,08	−0,03	−0,09	0,927
				Uni_2 vs. 3	35,65	12,79	1,01	2,79	0,008**
				number of semesters	6,60	3,15	0,57	2,10	0,042*
				school exp. (students)	0,05	0,20	0,03	0,24	0,811
				school exp. (mentors)	1,86	1,77	0,48	1,05	0,299
				university exp. (mentors)	0,08	1,06	0,01	0,08	0,937
Total	**.227**	**2,25**	**.055**	**Uni_1 vs. 3**	**26,26**	**21,71**	**0,42**	**1,21**	**0,233**
				Uni_2 vs. 3	23,98	10,26	0,83	2,34	0,024*
				number of semesters	4,872	2,524	0,515	1,93	0,060
				school exp. (students)	−0,01	0,16	−0,01	−0,06	0,954
				school exp. (mentors)	−0,15	1,42	−0,05	−0,10	0,919
				university exp. (mentors)	−0,81	0,85	−0,14	−0,95	0,345

Note.
* $p > .05$, ** $p > .01$, *** $p > .001$

Results show an effect of the university on the size of the difference between student and mentor ratings for Tasks and Verbal Input even when controlling for the other variables. Even when controlling for mentors' teaching experience, which significantly predicts Tasks, we still found that differences between students and mentors in Uni_1 are significantly larger than in Uni_3 and Uni_2. For Verbal Input, we found that differences between students and mentors are significantly larger in Uni_2 than in Uni_3 when controlling for the other variables (only the students' number of semesters is a significant predictor). The same seems to be valid for the Output Support and the Total Score, even though the regression model was not significant.

5. Discussion

5.1 Differences between student and mentor perceptions (RQ1) and universities (RQ2)

5.1.1 *Overall differences between students and mentors*

No significant differences were found in the two-factorial ANOVA (Table 4) between students and mentors across all scales (even though the value for Tasks approaches significance, see below), which is in line with previous studies by Bach (2013), Brodhäcker (2014), and Schubarth et al. (2012). Since we do not have a practicum control group that had not been introduced to the TIOS, we cannot determine whether this convergence is a training effect or would also be found with untrained raters. Yet, in another study on teachers' self-reports we find robust predictive effects of TIOS techniques on both L1 (Kersten et al., 2023) and L2 competences of their learners (Kersten et al., 2023, 2024).

5.1.2 *Overall differences between universities*

Between universities, however, there were significant rating differences for Verbal Input and Output Support, with a near-significant value for Tasks. As is obvious from Figure 1 (see conditional effects below), this difference seems to rely mainly on a pattern in which the student ratings of Uni_3 deviate systematically from the other two universities. Interestingly, these differences are not readily explainable based on structural differences such as degree programme (State Examination vs. Master of Education), practicum structure (Practicum Semester vs. Weekly Practicum), group size (individual vs. practicum in groups), or spacing of TIOS completion (after the lesson vs. after two consecutive lessons). If this were the case, the results obtained for Uni_2 rather than those for Uni_3 would have been most dissimilar, as Uni_1 and Uni_3 share the same basic structural features. We

thus hypothesise that factors beyond organisational surface structures, e.g., the contents of the respective university training, may have prepared students differently for their classroom teaching, so that their actual assessments differed, or that different types of awareness were instigated, resulting in different rating patterns.

5.1.3 *Overall differences in the interactions*

In addition, we found significant interactions between student/mentor ratings and universities for the Total Score, Tasks and Verbal Input, with scores for Output Support approaching significance. The near-significant values could be due to the small sample size and might become significant in a larger sample (see implications below). The interaction values indicate that the student/mentor ratings differ across universities.

To shed light on these differential effects, we examined the simple effects for all variables within all scales for the three universities (cf. Figure 1). The Total Score shows these overall differences most clearly. Here the students of Uni_3 differed from those of the other universities with significantly lower self-assessments, and from those of their own mentors. The same pattern is obvious in all subscales, as well (although not always statistically significant). However, there are also differential patterns in the simple effects, which become evident because the TIOS holds the advantage of detailed insights into the participants' perceptions due to its four subscales, as further discussed in 5.1.4.

5.1.4 *Simple effects to explain overall differences*

5.1.4.1 *Differences between students and mentors across universities*

The *t*-tests for Tasks and Verbal Input show significant but opposing differences between students and mentors. Regarding Tasks, the results showed that at Uni_2, the students assigned significantly higher self-ratings compared to their mentors' ratings, systematically assuming that their lessons fulfilled the pedagogical criteria of cognitive stimulation to a greater extent than the mentors. In line with previous studies finding over-positive self-perceptions by students (Bogard et al., 2017; Rothland, 2018), this suggests a more positive self-image of student teachers at Uni_2 regarding their language lessons' activities than their university mentors. In contrast, there is a significant difference at Uni_3 for Verbal Input, where student teachers rated the quality, adequacy and variation of their own teacher L2 input systematically lower than their mentors (Bodensohn & Schneider, 2008; Moser & Hascher, 2000). Whether this critical self-perception is indicative of a lack of self-confidence or rather a critical awareness of the student teachers' own classroom actions cannot be concluded from the data at hand, but seems a worthwhile endeavour for follow-up research. In any case, with these opposite patterns for Task Characteristics and Verbal Input we find instances of both students' under-

and overestimation within the same sample (cf. Bodensohn & Schneider, 2008; Rothland, 2018).

5.1.4.2 *Differences between students and students across universities*

Some conditional effects of differences between student groups could be explained against the backdrop of different content foci in the universities. For Task Characteristics, the student group from Uni_3 rated themselves significantly lower than in Uni_1. This might be attributed to the fact that at Uni_3, Task-Based Language Teaching is a major focus in the Master programme. This may have sensitised these students for task features such as meaningful communicative purposes or active problem-solving to a far greater extent than the other students, which could have led to a more critical self-evaluation that was not shared by their mentors.

For Verbal Input, Uni_2 stands out. While students at both Uni_1 and 2 have significantly higher ratings than the students at Uni_3, and students at Uni_2 have the highest input ratings in the sample, they still rate their performance slightly (though not significantly) lower than their mentors. This would be in line with that programme's specific focus on teacher language and input, coupled with the requirement for students to formulate their utterances in their lesson plans verbatim, including gestures and other scaffolds. This detailed preparation during planning presumably leads to a high degree of L2-use that is adapted to the learners' level, e.g., by highlighting with key words and phrases. The even higher (though not significantly so) ratings of their mentors (cf. Bodensohn & Schneider, 2008; Moser & Hascher, 2000) could be an indication that they are paying specific attention to this classroom interaction phenomenon which they prepared their students for.

5.1.5 *Possible explanation of overall differences*

In our data, Uni_3 consistently differs from both Uni_1 and Uni_2 on most measures. One reason might be that some of the teaching sessions at Uni_3 took place in grades 1/2 and 5/6, whereas Uni_1 and Uni_2 only included grades 3/4. This could, on the one hand, have led to different self-perceptions; on the other hand, the respective lessons require different types of scaffolding techniques geared at learners' varying levels of L2 competence and cognitive development (*contingency*, Kersten et al., 2024; van de Pol et al., 2010). Moreover, most mentors in the Uni_3 subsample hold secondary education degrees, which might result in greater benevolence (Bodensohn & Schneider, 2008; Moser & Hascher, 2000) when evaluating primary L2 lessons.

Another possible explanation for the differences between the subscales could be that students focus more intensely on preparing activities during the planning phase, in which they subsequently feel more confident. In comparison, it could be

more challenging to orient to their own teacher language, or they feel less confident regarding their own L2 use in class. Our findings thus suggest that students and mentors attach different importance to individual sub-aspects of lesson planning and delivery.

The differential observations for the subscales that the TIOS allows for thus underscore the importance of distinguishing between sub-aspects of L2 input quality, rather than applying global assessments of teaching competence (Bach, 2013; Brodhäcker, 2014; Festner et al., 2018; Schubarth et al., 2012).

5.2 Effects of teaching experience of students and mentors (RQ3 and RQ4)

The contribution of both the students' and mentors' pedagogic experience in the regression analyses explained 18.2% and 6.4% of the Total Score variance, respectively (Tables 5, 6). More specifically, students' number of semesters predicted Verbal Input, Output Support, and the Total Score negatively. In contrast to the participants in Pfingsthorn and Gehring (2015), our students appeared to become increasingly critical of their teaching techniques the longer they were at university, suggesting that a deeper engagement with language teaching methodology may have led to higher pedagogical content knowledge and greater awareness of its application in the classroom. Support for this conclusion comes from König et al. (2018), who found that students in higher semesters outperformed those in lower semesters regarding general pedagogical knowledge and pedagogical content knowledge. This observation is also in line with the Dunning-Kruger effect (Section 2.2). Moreover, the students' teaching experience did not predict their ratings, which suggests that the impact of knowledge imparted at university supersedes that of practical experience, at least for the pre-service teachers in our sample.

The same reasoning could apply to the mentors, whose results reveal the same pattern of negative predictive direction: The longer they had been teaching at university, the more critically they rated Verbal Input, while the length of school teaching experience negatively predicted the ratings for Output Support (Table 6).

Finally, as the linear regressions for RQ4 (Table 7) show, some differences between universities remained even when controlling for all four experience variables. While only the difference value (Δ) between Uni_2 and Uni_3 accounted for the Total Score variance, other differences in ratings between students and mentors along with experience predicted the scales differentially: Tasks were predicted by $\Delta_{U1,U3}$ and mentors' school experience, and Verbal Input and Output by $\Delta_{U2,U3}$ and students' semesters. These results again underscore the importance of differentially operationalising L2 input quality.

5.3 Limitations

Limitations of our study firstly concern the small sample size, especially with regard to our group of mentors, which reduces the generalisability and statistical power of the study, and the comparability of the three universities' programmes and practicum structure with regard to number, length, intensity and organisation of practicum. Operationalising prior teaching experience consistently across study programmes proved challenging, for instance, and more specific information about students' prior experiences and other possible confounding variables is warranted in future research. In addition, even though we followed a standardised procedure to introduce and practise the TIOS at all universities, students and particularly the mentors at Uni_1 might have been more familiar with the TIOS as the instrument was developed at this university. In light of these limitations, it is all the more striking that our data analyses revealed consistent and significant effects of teaching experience. While observations are, in general, preferable to self-reports, Kersten et al. (2023) showed that the TIOS can be used reliably for teachers' self-reports both in L1 and L2 contexts, which seems in line with the results presented here.

6. Conclusion

The results of our multi-site study showed differential ratings of students and mentors with regard to different dimensions of L2 input quality across student groups and universities. While students' self-perceptions and their mentors' assessments did not differ on the TIOS Total Score, a closer investigation of the underlying input dimensions (subscales) revealed both converging and diverging assessments, showing students' under- as well as over-estimations compared to their mentors. This provides a possible explanation of the heterogeneous evidence in previous research. Given the complex, multi-faceted nature of language teaching, this suggests, on the one hand, that various aspects of instructional quality are perceived differently by student teachers and more experienced observers, and on the other hand, that the TIOS subscales offer a valuable instrument for follow-up studies to examine these individual components more closely.

We identified the participants' pedagogic experience as key factors for these distributions, as well as differences between universities that are not explainable based on structural factors such as type of degree programme or practicum structure. We thus hypothesise that the programmes' different curricular foci predicted students' reflection skills and critical awareness differentially, although more follow-up research is needed to corroborate this claim based on systematic

information on programme contents What is more, the students' critical awareness of their teaching competences increased with their number of semesters at university, i.e., with their pedagogic experience, and this training was shown to have higher predictive power for their reflective competences than their actual practical experience. This finding is encouraging for primary language teacher education, as it indicates that the contents imparted in the university courses are indeed applied in the practicum and thus have great potential to enhance the competence of future primary English teachers.

Naturally, these assumptions need to be corroborated by future research including larger samples of students and mentors for more generalisability and eliciting more nuanced information on content foci for detailed results on programme effects. Still, our study offers added value for teacher training and L2 learning. In pointing out aspects in which trainees' self-assessments and external evaluations differ significantly, our findings can help improve pre-service primary English teacher education by identifying areas that should receive greater weight in lectures and preparatory courses for student teaching practicums. Additionally, the results might serve to enhance the quality of mentoring during teacher training. This, in turn, helps increase the quality of L2 teaching and learning, as teaching quality has been shown to affect learners' cognitive-linguistic competences (Praetorius et al., 2018, 2020), which is particularly important for L2 development. As our study has shown, the TIOS provides a valuable tool for primary language teacher education to capture both student teachers' and mentors' perceptions of teaching techniques. What is more, by asking both students and mentors to provide TIOS ratings, it can serve as a tool for both mentor feedback and self-evaluation. The detailed breakdown of instructional quality into the four subscales offered by the TIOS can lead to a deeper understanding and awareness of effective teaching strategies and thus help translate theories of ISLA into practical guidance for lesson planning and reflection in the young learners' language classroom.

Acknowledgments

We are highly grateful to all participants of the study, as well as schools and mentors, to Frederike Klose and Friederike Schirmag for data processing, and all student researchers in the STILE project I and II for the TIOS self-report version, notably Aenne Blackman, Fenja Funke, and Tooba Akram, and for data collection (www.uni-hildesheim.de/stile/; Kersten et al., 2023).

References

Bach, A. (2013). *Kompetenzentwicklung im Schulpraktikum: Ausmaß und zeitliche Stabilität von Lerneffekten hochschulischer Praxis*. Waxmann.

Bodensohn, R., & Schneider, C. (2008). Was nützen Praktika? Evaluation der Block-Praktika im Lehramt: Erträge und offene Fragen nach sechs Jahren. In R. Bodensohn (Ed.), *Lehrerbildung auf dem Prüfstand* (pp. 274–304). Verlag Empirische Pädagogik.

Bogard, T., Sableski, M.-K., Arnold, J.M. & Bowman, C.L. (2017). Minding the gap: Mentor and pre-service teachers' ability perceptions of content area literacy instruction. *Journal of the Scholarship of Teaching and Learning*, *17*(4), 44–66.

Bransford, J.D., Brown, A.L., & Cocking, R.R. (2000). How experts differ from novices. In J.D. Bransford, A.L. Brown & R.R. Cocking (Eds.), *How people learn: Brain, mind, experience, and school* (pp. 31–50). National Academy Press.

Brodhäcker, S. (2014). *Unterrichtsplanungskompetenz im Praktikum: Einflussfaktoren auf die Veränderung der wahrgenommenen Kompetenz von Studierenden*. Waxmann.

Dassa, L., & Nichols, B. (2019). Self-efficacy or overconfidence? Comparing preservice teacher self-perceptions of their content knowledge and teaching abilities to the perceptions of their supervisors. *The New Educator*, *15*(2), 156–174.

Dunning, D. (2011). The Dunning–Kruger effect: On being ignorant of one's own ignorance. *Advances in Experimental Social Psychology*, *44*, 247–296.

Ellis, R., & Shintani, N. (2014). *Exploring language pedagogy through second language acquisition research*. Routledge.

Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2020). *Task-based language teaching: Theory and practice*. Cambridge University Press.

Festner, D., Schaper, N., & Gröschner, A. (2018). Einschätzung der Unterrichtskompetenz und -qualität im Praxissemester. In J. König, M. Rothland & N. Schaper (Eds.), *Learning to practice, learning to reflect? Ergebnisse aus der Längsschnittstudie LtP zur Nutzung und Wirkung des Praxissemesters in der Lehrerbildung* (pp. 163–193). Springer VS.

Gass, S.M., Behney, J., & Plonsky, L. (2020). *Second language acquisition: An introductory course (5th ed.)*. Routledge.

Gröschner, A., Schmitt, C., & Seidel, T. (2013). Veränderung subjektiver Kompetenzeinschätzungen von Lehramtsstudierenden im Praxissemester. *Zeitschrift für Pädagogische Psychologie*, *27*(1–2), 77–86.

Helmke, T., Helmke, A., Schrader, F.-W., Wagner, W., Nold, G., & Schröder, K. (2008). Die Videostudie des Englischunterrichts. In DESI-Konsortium (Eds.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch: Ergebnisse der DESI-Studie* (pp. 345–363). Beltz.

Hogan, T., Rabinowitz, M., & Craven III, J.A. (2003). Representation in teaching: Inferences from research of expert and novice teachers. *Educational Psychologist*, *38*(4), 235–247.

Kersten, K. (2020). Der Zusammenhang von kognitiven Fähigkeiten und Zweisprachigkeit im frühen L2-Erwerb: Individuelle und externe Faktoren. In H. Böttger, J. Festman & T. Müller (Eds.), *Language education & acquisition research: Blick auf den frühen Spracherwerb* (pp. 81–118). Klinkhardt.

Kersten, K. (2021). L2 input and characteristics of instructional techniques in early foreign language classrooms: Underlying theory and pedagogical practice. *The European Journal of Applied Linguistics and TEFL*, *10*(2), 27–59.

Kersten, K. (2023). The Proximity of Stimulation Hypothesis: Investigating the interplay of social and instructional variables with the cognitive-linguistic skills of young L2 learners. In K. Kersten & A. Winsler (Eds.), *Understanding variability in second language acquisition, bilingualism, and cognition: A multi-layered perspective* (pp. 131–159). Routledge.

Kersten, K., Blackman, A., Funke, F. L., Akram, T., Kliebisch, M., & Koch, M. J. (2023). Input quality affects L1 and L2 proficiency and moderates the effect of socioeconomic status. In H. Böttger & N. Schlüter (Eds.), *Fortschritte im frühen Fremdsprachenlernen. Band zur 5. Konferenz* (pp. 255–280).

Kersten, K., Bruhn, A.-C., Ponto, K., Böhnke, J., & Greve, W. (2018). Teacher Input Observation Scheme (TIOS). *Studies on Multilingualism in Language Education*, *4*, Hildesheim University.

Kersten, K., Bruhn, A.-C., Ponto, K., Koch, M. J., & Greve, W. (2024). Assessing L2 input using the Teacher Input Observation Scheme (TIOS): Input predicts L2 acquisition of young learners differentially [Manuscript submitted for publication]. Mannheim University.

Kersten, K., Schelletter, C., Bruhn, A.-C., & Ponto, K. (2021a). Quality of L2 input and cognitive skills predict L2 grammar comprehension in instructed SLA independently. *Languages*, *6*(3), 124–142.

Kersten, K., Bruhn, A.-C., Ponto, K., Böhnke, J., & Greve, W. (2021b). Teacher Input Observation Scheme und Manual — Deutsche Version für L2. *Studies on Multilingualism in Language Education*, *7*, Hildesheim University.

Klingebiel, F., Mähler, M., & Kuhn, H. P. (2020). Was bleibt? Die Entwicklung der subjektiven Kompetenzeinschätzung im Schulpraktikum und darüber hinaus. In I. Ulrich & A. Gröschner (Eds.), *Praxissemester im Lehramtsstudium in Deutschland: Wirkungen auf Studierende*. Edition ZfE (Vol. 9, pp. 179–207). Springer.

König, J., Doll, J., Buchholtz, N., Förster, S., Kaspar, K., Rühl, A.-M., Strauß, S., Bremerich-Vos, A., & Kaiser, G. (2018). Pädagogisches Wissen versus fachdidaktisches Wissen? Struktur des professionellen Wissens bei angehenden Deutsch-, Englisch- und Mathematiklehrkräften im Studium. *Zeitschrift für Erziehungswissenschaft*, *21*(3), 1–38.

Kormos, J. (2011). Speech production and the Cognition Hypothesis. In P. Robinson (Ed.), *Second language task complexity* (pp. 39–60). John Benjamins.

Leow, R. P. (2018). *Explicit learning in the L2 classroom. A student-centred approach*. Routledge.

Leow, R. P. (2018). Explicit learning and depth of processing in the instructed setting: Theory, research, and practice. *Studies in English Education*, *23*(4), 769–801.

Li, P., Jeong, H. (2020). The social brain of language: Grounding second language learning in social interaction. *npj Science of Learning*, *5*(8), 1–9.

Loewen, S. (2020). *Introduction to instructed second language acquisition*. Routledge.

Loewen, S., Sato, M. (2018). Interaction and instructed second language acquisition. *Language Teaching*, *51*(3), 285–329.

Long, M. H. (2015). *Second language acquisition and task-based language teaching*. Wiley-Blackwell.

Moser, P. & Hascher, T. (2000). *Lernen im Praktikum. Projektbericht*. Forschungsstelle für Schulpädagogik und Fachdidaktik. https://edudoc.ch/record/2718?ln=de

Muñoz, C. (2019). A new look at age: Young and old L2 learners. In J. W. Schwieter & A. Benati (Eds), *The Cambridge handbook of language learning* (pp. 430–450). Cambridge University Press.

Nilsen, T., Scherer, R., Gustafsson, J.-E., Teig, N. & Kaarstein, H. (2020). Teachers' role in enhancing equity — A multilevel structural equation modelling with mediated moderation. In T. S. Frønes, A. Pettersen, J. Radišić & N. Buchholz (Eds.), *Equity, equality and diversity in the Nordic model of education* (pp. 173–196). Springer.

Pfingsthorn, J. & Gehring, W. (2015). Ready to teach: The development of teacher trainees' perceived competence. In Pixel (Ed.), *Conference Proceedings of the 8th International Conference ICT for Language Learning* (pp. 336–342). Libreriauniversitaria.it. https://conference.pixel-online.net/ICT4LL/files/ict4ll/ed0008/FP/1973-LTT1236-FP-ICT4LL8.pdf

Praetorius, A. K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of Three Basic Dimensions. *ZDM*, *50*, 407–426.

Praetorius, A.-K., Klieme, E., Kleickmann, T., Brunner, E., Lindmeier, A., Taut, S., & Charalambous, C. (2020). Towards developing a theory of generic teaching quality. Origin, current status, and necessary next steps regarding the Three Basic Dimensions Model. In A. Praetorius, J. Grünkorn, & E. Klieme (Eds.), *Empirische Forschung zu Unterrichtsqualität. Theoretische Grundfragen und quantitative Modellierungen* (pp. 15–36). Beltz Juventa.

Rothland, M. (2018). Yes, we can! Anmerkungen zur trügerischen "Kompetenzentwicklung" von Lehramtsstudierenden im Praxissemester. *Beiträge zur Lehrerinnen- und Lehrerbildung*, *36*(3), 482–495.

Schempp, P., Tan, S., Manross, D. & Fincher, M. (1998). Differences in novice and competent teachers' knowledge. *Teachers and Teaching: Theory and Practice*, *4*(1), 9–20.

Schubarth, W., Speck, K., Seidel, A., Gottmann, C., Kamm, C. & Krohn, M. (2012). Das Praxissemester im Lehramt — Ein Erfolgsmodell? Zur Wirksamkeit des Praxissemesters im Land Brandenburg. In W. Schubarth, K. Speck, A. Seidel, C. Gottmann, C. Kamm, & M. Krohn (Eds.). *Studium nach Bologna: Praxisbezüge stärken?! Praktika als Brücke zwischen Hochschule und Arbeitsmarkt.* (pp. 137–169). Springer VS.

Schüle, C., Besa, K.-S., Schrick, J. & Arnold, K.-H. (2017). Die Veränderung der Lehrerselbstwirksamkeitsüberzeugung in Schulpraktika. *Zeitschrift für Bildungsforschung*, *7*(1), 23–40.

Spada, N. (2019). Classroom observation research. In J. W. Schwieter, & A. Benati (Eds.), *The Cambridge handbook of language learning* (pp. 186–207). Cambridge University Press.

Spada, N. (2024). *The COLT Observation Scheme: Digital versions and updated research applications*. John Benjamins.

Spada, N., Fröhlich, M. (1995). *COLT — Communicative Orientation of Language Teaching Observation Scheme: Coding conventions and applications*. National Centre for English Language Teaching and Research.

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471–483). Routledge.

Truscott, J., & Sharwood Smith, M. (2019). *The internal context of bilingual processing*. John Benjamins.

Tsui, A. B. M. (2009). Distinctive qualities of expert teachers. *Teachers and Teaching: Theory and Practice*, *15*(4), 421–439.

Weitz, M. (2015). *Die Rolle des L2-Inputs in bilingualen Kindergärten*. Peter Lang.

Weitz, M., Pahl, S., Flyman Mattsson, A., Buyl, A. & Kalbe, E. (2010). The input quality observation scheme (IQOS): The nature of L2 input and its influence on L2 development in bilingual preschools. In K. Kersten, A. Rohde, C. Schelletter, & A. K. Steinlen (Eds.), *Bilingual preschools (Volume I): Learning and development* (pp. 5–44). WVT.

Wilden, E. & Porsch, R. (2019). The impact of teaching quality and learning time on primary EFL learners' receptive proficiency: Preliminary findings from the TEPS study. *AILA Review*, *32*(1), 160–177.

Wilden, E., Porsch, R. & Schurig, M. (2020). An early start in primary EFL education and the role of teacher qualification and teaching quality. *Language Teaching for Young Learners*, *2*(1), 28–51.

Wolff, C. E., Jarodzka, H., van den Bogert, N., & Boshuizen, H. P. A. (2016). Teacher vision: Expert and novice teachers' perception of problematic classroom management scenes. *Instructional Science*, *44*(3), 243–265.

Wolff, C. E., Jarodzka, H., & Boshuizen, H. P. A. (2021). Classroom Management scripts: A theoretical model contrasting expert and novice teachers' knowledge and awareness of classroom events. *Educational Psychologist*, *38*(4), 131–148.