

Using evaluation to promote change in language teacher practice

Rosemary Erlam
The University of Auckland

Recent literature in teacher education has argued for a shift away from the development of teacher cognitions as a goal of teacher education, to the development of core practices which would make a difference to students' lives in the classroom (Ball & Forzani, 2009; Kubanyiova & Feryok, 2015; Zeichner, 2012). Hiebert and Morris (2012) propose that these key practices would be embedded into instructional contexts and preserved as lesson plans and as common assessments.

This paper focuses on the evaluation tools developed for an in-service professional development programme for language teachers (the Teacher Professional Development Languages (TPDL) programme: <http://www.tpd.ac.nz/>). TPDL is a year-long programme for teachers of foreign languages in NZ schools. Programme participants are visited by TPDL In-School support facilitators four times during the course of the year. The facilitators observe their teaching practice and then use two key documents, the 'Evidence of Principles and Strategies (EPS) portfolio' and the 'Progress Standards' to assist teachers to evaluate their practice against key criteria. As the year progresses the teachers are increasingly encouraged to take ownership and control of the use of these tools, so that by Visit 4, the evaluation is conducted as a self-assessment. This paper evaluates these tools and considers evidence for their validity. Data is presented from the case study of one teacher, to further demonstrate how the tools are used and to document evidence for any change in teaching practice.

Key words: Language teacher education, validation, foreign language teaching, self-assessment

Background

Recent literature in teacher education documents a shift or turn in emphasis from the development of teacher cognitions to the development of core practices (Zeichner, 2012). Hiebert and Morris (2012) explain that a common approach to educating teachers works from the assumption that improving their knowledge and skills improves their teaching. They suggest a reversal of this relationship and claim that creating opportunities for teachers to study, test and revise instructional methods will lead to improved skills for teaching. Ball and Forzani (2009, p. 497) claim that teacher education needs to pay attention to the core tasks that teachers must execute in order for their pupils to learn. This shift in emphasis is also present in language teacher education literature; Kubaniyiova & Feryok (2015) stress the need for an emphasis on those practices which would make a difference to the lives of pupils in the classroom and so impact positively on student learning.

Once such a set of practices were identified, it would be necessary for teachers to have the opportunity to both engage in these practices and to 'measure their performance against exemplars' (Ball & Forzani, 2009, p. 499). Ideally these opportunities would be available to teachers in their own teaching contexts, so that they could get feedback about whether or not the practices and methods of teaching would be effective in these contexts (Ball & Forzani, 2009). Hiebert & Morris (2012) suggest that there would be an ongoing need to test and improve any set of core practices and thus stress the role of assessment in measuring the effectiveness of lessons and the nature of student outcomes. They suggest that lesson plans and common assessments would form the important function of documenting and recording the learning and knowledge that would result from these processes.

It would seem then, that with this new emphasis on 'the work of teaching' and on the provision of opportunities for teachers to evaluate their performance in relation to core practices, that there is a role for self-assessment in teacher education. Ross and Bruce (2007) document the literature that has demonstrated that teaching students to self-assess accurately leads to higher achievement (e.g. McDonald & Boud, 2003; Ross, Rolheiser, & Hogaboam-Gray, 1999; Ross, Hogaboam-Gray & Rolheiser, 2002), but point out that self-assessment has been a neglected strategy for facilitating the professional growth of teachers. In their own study they demonstrated that a self-assessment tool contributed to the professional growth of a grade 8 mathematics teacher, by helping him to understand and recognise excellence in teaching and to identify goals and gaps between preferred and actual practices.

Within the language teacher education literature Freeman, McBee Orzulak and Morrissey (2009) claim that the shift in teacher education from an emphasis on the end product to one on the teaching process, entails a central role for self-assessment.

Scarino (2014) stresses the importance of getting the teacher to analyse and reflect on their own performance. She claims that the best way to empower the teacher to evaluate their own teaching practice is to create opportunity for dialogue and interaction between teacher educators and teachers.

To have confidence in the use of any assessment tool, that is, to be able to support the inferences that are made concerning the ability(ies) measured, it is important to be able to argue that it has validity (Weir, 2005). Weir (2005) claims that evidence for validity of an instrument can be collected prior to its use (a priori) or after its use (a posteriori). One type of a priori evidence is theory-based validity, where it can be demonstrated that the tool is informed by a well theorised and defensible understanding of the behaviour or ability that is being assessed. Context validity, also a focus at this stage, concerns the fit that there is between the assessment and its administrative setting or context. Scoring validity, an important aspect of the a posteriori validation procedure, considers the extent to which assessment results are stable and consistent over time, as well as free from bias. A further component, that is consequential validity, where the potential and actual social consequences of the use of an assessment tool are evaluated, Weir attributes to Messick (1989). A related notion is the idea of backwash, where the effect of assessment on teaching and also on learning is evaluated (Hughes, 2003; Shohamy, 2001).

This paper describes two evaluation tools developed for the Teacher Professional Development Languages (TPDL) programme, an in-service professional development programme for language teachers, and used with the aim of promoting change in teacher practice. The first is referred to as the Evidence of Principles and Strategies (EPS) portfolio and is used in what is called a 'learning' discussion with a teacher educator, to help teachers evaluate their own teaching practice. The second document, referred to as 'The Progress Standards', evaluates teacher practice against three key criteria. These tools are thus what Hiebert and Morris (2012, p. 93) would refer to as 'artefact[s], or knowledge product[s]' in that they enable knowledge, in this case, about how to effect teacher learning and change in teacher practice, to be shared.

The paper will first give a brief overview of the programme. It will then describe these evaluation tools and how they are used within the programme, at the same time presenting and discussing a range of evidence for their validity. Data investigating the effectiveness of the tools in effecting changes in teaching practice that impact on student learning will also be presented and discussed.

The TPDL (Teacher professional development languages) programme

TPDL began in 2005 as a Ministry of Education funded project. It followed on from a recommendation that all students in Years 7 to 10 in New Zealand schools have the opportunity to learn a language other than English and from a report highlighting the lack of professional development opportunities for language teachers (Gibbs & Holt, 2003). TPDL caters for both teachers who have limited or no language teaching experience as well as those who might have considerable experience teaching a language other than English. It caters for teachers of ten languages other than English: French, Chinese, Japanese, Spanish, German, Samoan, Tongan, Cook Islands Maori, Tokelauan and Niuean. While initially designed to target teachers of students in Year 7 to 10 (i.e., students aged 11 to 14 years), it is now open to those teaching at all levels of the school curriculum. The TPDL programme aims to develop teachers' language proficiency (teachers already fluent in one language are encouraged to learn another, so as to have the experience of being a beginner language learner again) and to give them the pedagogical skills necessary to teach another language effectively so that they might improve student learning outcomes. There are three interrelated components of the programme. Firstly, participants enrol in language study courses. As part of this component they have the chance to sit international language exams. Secondly, participants undertake the study of second language acquisition (SLA) pedagogy. This part of the program is delivered over a total of eight days, in four blocks, and includes a University of Auckland Level 3 course. The third component of the programme is referred to as 'in-school support' (Insley & Thomson, 2008) and allows for participants to be visited four times over the course of the year by an In-School Support Facilitator (ISSF). The ISSF observes the teacher teaching a language lesson in their own teaching context and writes down as many teacher, or student, utterances (depending on which visit in the sequence of four they are observing) as they can during the lesson. Following this lesson the ISSF and teacher use this document of teacher/student utterances, the EPS and the Progress Standards to discuss and evaluate together the lesson that has just been observed. The author of this paper has been the Academic Director of TPDL since 2010. In this role she takes responsibility for the overall quality of the programme, with a particular focus on the taught pedagogy component. She did not design the evaluation tools but has been involved in some of the ongoing discussions around their use and modification.

The theoretical basis of TPDL

The turn, in teacher education, from an emphasis on teacher cognitions and knowledge to one on core classroom practices has already been discussed. An obvious challenge is to identify the core practices that would be worthy of focus in

teacher education programmes. For second/foreign language teaching it would seem to be important to start by developing a coherent theoretical perspective on how (a) language(s) is/are learnt. Given the enormous body of research on second language acquisition, this is not an easy task, nor once this were achieved would it be easy to extrapolate a set of practices that would account for effective language teaching (Ellis, 2005a). Indeed, it needs to be acknowledged that there are competing theories and different perspectives on how a second or foreign language might be most effectively taught.

The TPDL programme has been heavily influenced by research and literature in the area of second language acquisition. Key findings, referred to in the programme as 'principles', have determined the criteria against which teaching practice is evaluated and the content of the pedagogy programme. This paper will restrict itself primarily to a discussion of the former, given that it is more particularly concerned with the evaluation process.

In 2005 Ellis was asked by the Ministry of Education to conduct a literature review with the aim of generating a set of research-based principles that could provide teachers with a guide to effective language teaching and serve as a basis against which their own teaching could be evaluated. The resulting document outlined 10 principles derived from a variety of theoretical perspectives that provided a psycholinguistic account of language learning. Ellis was careful to note that other perspectives, such as educational theories of 'good teaching' were not included (Erlam & Sakui, 2006, p. 7). He was also careful to point out that these could only be 'provisional' rather than definitive specifications as to what might underline effective teaching practice and that they would need to be tried out by teachers in their different teaching contexts. With respect to TPDL it was particularly crucial to recognise that 'one size may not fit all' given the variety of languages that the programme caters for along with the variety of teaching contexts. The 10 principles are listed below (Ellis, 2005a, 2005b).

- Principle 1** Instruction needs to ensure that learners develop both a rich repertoire of formulaic expressions and a rule-based competence
- Principle 2** Instruction needs to ensure that learners focus predominantly on meaning
- Principle 3** Instruction needs to ensure that learners also focus on form
- Principle 4** Instruction needs to be predominantly directed at developing implicit knowledge of L2 while not neglecting explicit knowledge
- Principle 5** Instruction needs to take into account the learner's "built-in syllabus"
- Principle 6** Successful instructed language learning requires extensive L2 input
- Principle 7** Successful instructed language learning also requires opportunities for output

- Principle 8** The opportunity to interact in the L2 is central to developing L2 proficiency
- Principle 9** Instruction needs to take account of individual differences in learners
- Principle 10** In assessing learners' L2 proficiency it is important to examine free as well as controlled production

In view of the fact that, firstly, Ellis's 10 principles did not discuss the role of cultural knowledge in language teaching and learning and, secondly, in consideration of the importance accorded to cultural knowledge in the 'learning languages' learning area of the New Zealand curriculum (Ministry of Education, 2007) the Ministry of Education commissioned a second report. The resulting document, known as the Newton report, was published in 2010 (Rivers, 2010). It also contained an extensive review of the literature and outlined 6 principles of Intercultural communicative language teaching. These are listed below:

- iCLT Principle 1** ICLT integrates language and culture from the beginning
- iCLT Principle 2** ICLT engages learners in genuine social interaction
- iCLT Principle 3** ICLT encourages and develops an exploratory and reflective approach to culture and culture-in-language
- iCLT Principle 4** ICLT fosters explicit comparisons and connections between languages and cultures
- iCLT Principle 5** ICLT acknowledges and responds appropriately to diverse learners and learning contexts
- iCLT Principle 6** ICLT emphasises intercultural communicative competence rather than native-speaker competence.

From theory to practice: the EPS evaluation tool

In conjunction with the initiative that resulted in the writing of the 10 'Ellis' principles, the Ministry also funded a project that looked for evidence of these principles in classroom practice in New Zealand schools. Erlam and Sakui (2006) therefore conducted a series of case studies in which they documented evidence of the 10 principles in French and Japanese classrooms. The associated advantage of this project was that the principles needed to be 'operationalised' as a series of 'behaviours', in other words, the essential skills that the principles outlined had to be translated into classroom practice. Erlam and Sakui (2006) documented, therefore, in an observation schedule, the behaviours that they would look for in the classroom as evidence of each of the principles. In recognition of the fact that they were able to observe only 4 lessons for each of the teachers in the project, they also designed interview questions that would more completely inform them about classroom practice. The publication that ensued from this project, that is, the case studies,

which included the observation schedule and interview questions, was sent to all schools in New Zealand.

The EPS drew heavily on this document. This can be demonstrated by taking the first page of the EPS as displayed below in Table 1. The first two items of this document draw on Principles 1 (Instruction needs to ensure that learners develop both a rich repertoire of formulaic expressions and a rule based competence) and 6 (Successful instructed language learning requires extensive L2 input). Both of these first two items on the EPS were adapted from Erlam and Sakui (2006).

Table 1. An extract from p. 1 & 2 of the EPS

Evidence of Principles and Strategies 2014					
<i>Evidence of teachers demonstrating principles, strategies and resources that are effective in improving student achievement in the target language.</i>					
1. Formulaic chunks in the TL that students were heard using:					
Visit 2:					
Visit 3:					
Visit 4:					
2. In the following ways you provide TL input for your students:					
Using TL for Classroom Management					
Visit 2	5	4	3	2	1
Visit 3	5	4	3	2	1
Visit 4	5	4	3	2	1
Using TL for Social Goals					
Visit 2	5	4	3	2	1
Visit 3	5	4	3	2	1
Visit 4	5	4	3	2	1
Using TL for Language Goals					
Visit 2	5	4	3	2	1
Visit 3	5	4	3	2	1
Visit 4	5	4	3	2	1
<i>(5 being entirely in TL and 1 being entirely in English)</i>					
Other sources of TL input during the lesson					
Visit 2:					
Visit 3:					
Visit 4:					

Participants in the TPDL programme are informed that the 10 Ellis principles form the pedagogical base of the programme and are given a copy of the literature review containing these. The EPS document is not used during Visit 1, rather the 'learning discussion' that takes place at the end of this visit uses a shorter one page document which is, as noted on the document, 'based on "Principles of Effective Instructed Language Learning"'. By the time that Visit 2 with the ISSF takes place, all participants will have attended Days 1-2 of the pedagogy component of the programme which aim to help them understand how the principles might be embodied in classroom practice. In the remaining 6 days of the pedagogy component participants are introduced to the theoretical rationale for the 10 principles. They also complete, putting themselves in the role of language learners, a variety of language tasks, some of which are designed in the language they are teaching and some of which are in a language with which they are most likely not familiar (e.g. Te Reo Māori); they subsequently discuss together to what extent these tasks exemplify classroom practice consistent with the principles and how applicable they would be to their own teaching contexts. These pedagogy days also address topics such as Task-based language teaching and the New Zealand curriculum. The iCLT principles are normally introduced during the second half of the pedagogy component. Opportunities are also provided for participants to work in groups with others according to the language they teach and also, at times, according to their teaching contexts, in recognition of the fact that they come from very different backgrounds with a wide range of experience.

From theory to practice: The Progress Standards

As has already been described, the EPS documents evidence in the classroom for the Ellis principles. The Progress Standards document is used subsequently to rate teacher practice against three key criteria (see Appendix 1 for this document). These are: teacher use of the target language, student use of the target language and opportunities for student interaction. These were chosen based on the knowledge that the teacher educators (Insley & Thomson, 2008) had gained over time of those aspects of teacher practice that they considered not to be in evidence in language classrooms and which were most crucial for beginner language learners.

In conclusion, the fact that both these tools are informed by research that has investigated theoretical understanding of second language acquisition processes and effective language teaching pedagogy, constitutes evidence of theory-based validity. It could be argued, of course, that not all language teaching theorists might agree with the 10 principles underlying the tools, notwithstanding, however, as Erlam (2008) documents, the 'uptake' of these principles in the New Zealand context and the impact on language teacher education has been considerable. Part of the reason

for this is that the research which informed the design of the tools had been conducted in the New Zealand classroom context, a fact which is, arguably, further evidence of theory-based validity (Erlam and Sakui, 2006).

Adapting evaluation to context

The EPS evaluation tool has been modified and adapted, often over time and following iterations of its use, to be more helpful in its use with TPDL participants. For example, while the first versions of the EPS did acknowledge the intercultural dimension of language learning and referenced Byram (1997), the EPS was redesigned following the publication of the Newton report (2010) to incorporate specific reference to the set of iCLT principles that he identifies.

There have also been ongoing changes to the EPS document as it continues to serve the needs of the contract. In 2008 the Ministry of Education documented the disparity between the literacy results of Pasifika and European students. This led to a focus on ensuring learning gains for students who might be underachieving in different areas of the curriculum. There was therefore a modification to the Item of the EPS that looked for evidence of Ellis principle 9 and which had originally required teachers to consider:

‘In these ways you catered for individual differences’ (extract from EPS, 2010 version). This became (Table 2):

Table 2. Extract from EPS, 2011 version, item 7

7. In these ways you catered for **individual differences** and, in addition, addressed the learning needs of Māori and Pasifika students:

Visit 2

Visit 3

Visit 4

The TPDL evaluation tools have therefore been adapted from the document that had informed their design (Erlam and Sakui, 2006) to be more appropriate and user friendly in the New Zealand context. These changes and modifications which continue to ensure that the tools are appropriate for the settings in which they are used constitute some evidence for context validity (Weir, 2005).

Empowering teachers to self-evaluate

A handbook written for In-School Facilitators contains specific instructions for how the EPS is to be used with teachers following classroom observation in Visits 2 to 4. As discussed previously the EPS is used alongside a document recording 'teacher utterances and student utterances' (Visit 2) and a document recording the 'utterances of two students chosen by the ISSF' (Visits 3 & 4). ISSFs are told in the guidelines that teachers are to be encouraged to take an 'inquiry' approach as they look at the document of teacher/student utterances and the EPS. They are told that they should initially avoid any subjective discussion or comment on the lesson, but rather start by examining, with the teacher, the evidence from the two documents, using it to inform the evaluation and discussion. The aim is to help the teacher make connections between the research that they have been exposed to and their classroom practice, and as they do so, gather evidence of their teaching practice that might support student learning, and, increasingly throughout the programme, to document any evidence of student learning. In using the EPS, the ISSF must increasingly aim to empower the teacher to take ownership of the discussion (Scarino, 2014) and to 'express the presence or absence of the evidence for the principles him/herself' (TPDL, 2014, p. 5). It is noted that this is increasingly important in Visits 3 and 4 and that there is, in Visit 4, the expectation that the teacher has total responsibility for leading the discussion.

In the information given to the ISSFs, care is taken to ensure that the discussion with the teacher around the evidence provided by the EPS and the document of teacher/student utterances, is not seen as a 'test' that teachers have to pass. The very term used to designate the discussion between the ISSF and the teacher, that is, 'learning conversation', takes the emphasis away from evaluation and emphasizes the importance of supporting the teacher to grow in their reflection and understanding (Mathew & Poehner, 2014).

However, it would not be right to conclude that there are no expectations with respect to teachers' progress or performance during their time on the TPDL programme. Indeed, as has already been explained, after each of Visits 2 to 4, teachers' progress is rated as 'expected' or 'accelerated' against three key criteria on the Progress Standards document (see Appendix 1). Once again though, teachers are encouraged to self-rate their progress against these standards. According to the Director of the programme, in most cases, because of the evidence that has been gathered during the EPS discussion, teachers are completely accurate in their self-assessment and their self-rating concurs with that of the ISSF.

The ISSFs are reminded in their handbook that the EPS standards have been developed by the team and not by the Ministry and that if they are not 'working'

they will need to be refined. Because there is no pressure from the Ministry of Education for teachers to meet the criteria and standards that have been developed by the TPDL team, changes to the EPS that have been modified to suit programme requirements (as documented earlier) have primarily been motivated by the TPDL team. In discussion with the Director of the programme, the crucial determinant of the success of the evaluation process that is conducted with the teacher (and, indeed, the determinant of the success of the whole TPDL programme) is whether or not teachers learn to be able to self-reflect on their own teaching. The ultimate aim is that teachers will be empowered to plan lessons where they can stand back and gather evidence of the effectiveness of their own teaching.

In conclusion then, there is some, although rather anecdotal, evidence that in using the EPS and Progress Standards documents over time teachers develop greater skill and accuracy in evaluating their own practice. More principled and rigorous investigation is needed of the process by which and the extent to which teachers are empowered to accurately self-evaluate and reflect on their own classroom practice. This would constitute some evidence for the consequential validity (Weir, 2005) of the evaluation tools and would allow for comparison with previous research which has documented growth in teacher ability to recognise and set goals for improvement (Ross and Bruce, 2007).

Consistency in use of the evaluation tools

There is ongoing moderation of ISSFs as they conduct 'learning discussions' with TPDL participants. Each term each ISSF is accompanied by another ISSF during one visit. The moderator will gather evidence to help the ISSF establish whether they effectively assisted the teacher to reflect on and inquire into their own teaching. In this way there is also an emphasis on the reliability of the EPS/learning discussion, given that there is a team of ISSFs and therefore a need to establish some consistency in the way the evaluation process is implemented. These processes constitute some evidence for the ongoing scoring validity of the tools (Weir, 2005). Further research could investigate the extent to which teachers' experience is consistent across variation in ISSF personnel and teaching context. The whole issue of reliability and scoring validity is relevant to the reporting of programme outcomes, as will be discussed below.

Impact on teaching practice

A Ministry of Education requirement is the ongoing reporting of programme outcomes, along with evidence to support claims. The researcher, as the Academic

Director of TPD, is involved in documenting evidence of programme effectiveness at six-monthly intervals, under different categories, one of these being: 'changes in teaching practice that improve learning.' In order to report evidence for this criterion, she is reliant on data collected by the ISSFs. In June 2014 she reported as follows:

' . . . indication of change in teaching practice is the increased use of pair and group work in the classroom. . . .A number of teachers were already using pair and group work in their language classrooms during Visit 1. However, there were eight teachers who did not use group/pair work during Visit 1 but had incorporated this into Visit 2 classrooms.'

In December 2014, she documented evidence of 'change in teaching practice that impacted on learning' as follows:

'Ninety two percent of teachers showed an increase in the amount of target language they used in the classroom between Visits 1 and 4. This increase in the teacher use of the target language in the classroom, along with increased activities that teachers provided for students to use and interact in the oral language, had a considerable impact on student learning. This is evidenced by the fact that at Visit 4, in 84% of classrooms, students were using language that was quantitatively and qualitatively superior to that used in Visit 1'.

This data, while suitable for Ministry reporting purposes, cannot be used for research purposes because it does not satisfactorily fulfil requirements for reliability. It is collected by different ISSFs, with no independent verification, so subject to variability. These ISSFs cannot, and indeed, do not aim to, record all of the classroom discourse of the lesson they observe, so it is very likely that their observations are incomplete. This demonstrates how the evaluation tools that are reported on in this paper, the EPS and the Progress Standards, are restricted in terms of use to the context for which they were designed. It was with this in mind that the researcher undertook a research project aiming to conduct an independent investigation of programme outcomes. She also wanted to track the learning journey of participants in the programme.

Impact on teaching practice: a case study

The researcher therefore invited all participants in the programme in 2014 to agree to be part of a research project. Participation in the project involved agreeing to the researcher observing and audio-recording a lesson prior to the teacher's participation in the TPD programme and conducting an interview. The researcher also observed and recorded all lessons during ISSF visits and conducted a final

interview at the end of the programme. Students were invited to be part of the study by agreeing to be audio-recorded as they took part in lessons. Regrettably, in retrospect, the data collection process did not include audio-recording of the ISSF-teacher discussion as they completed the evaluation tools together.

Of the 65 participants in TPDL in 2014, unfortunately only two agreed to be part of this project, and only one indicated their willingness to participate prior to the commencement of ISSF Visit 1. This participant will be referred to as Jane for the purposes of anonymity. Jane is a UK trained teacher of French with a NZ postgraduate qualification (PGDip in Primary Teaching, Massey University). Jane has over 20 years of experience teaching French and at the time of the study was teaching Year 5 students (aged approximately 9/10 years) at a private girls' school. The students in Jane's class received just one 50 minute lesson of French a week.

Data will be presented in terms of evidence from Jane's classroom of her performance in relation to the three key criteria highlighted on the Progress Standards document. These are:

1. teacher use of the target language (TL).
2. student use of the TL.
3. provision of opportunities for students to interact in the target language and use it as a tool for meaningful communication

Data collected by the ISSF in relation to these three criteria will be compared with data collected by the researcher.

Teacher use of target language

Figures 1 and 2 document the ISSF's rating of the teacher's use of the target language in Visit 1 (Figure 1) and Visits 2 to 4 (Figure 2).

TL input by the teacher. To what extent is the TL used by the teacher for: (5 being entirely in TL and 1 being entirely in English)

Classroom management:	5	4	3	2	1
Social goals:	5	4	3	2	1
Language goals:	5	4	3	2	1

Figure 1. Teacher use of the TL as recorded on 'Visit 1 reflection' document

2. In the following ways you provide TL input for your students:

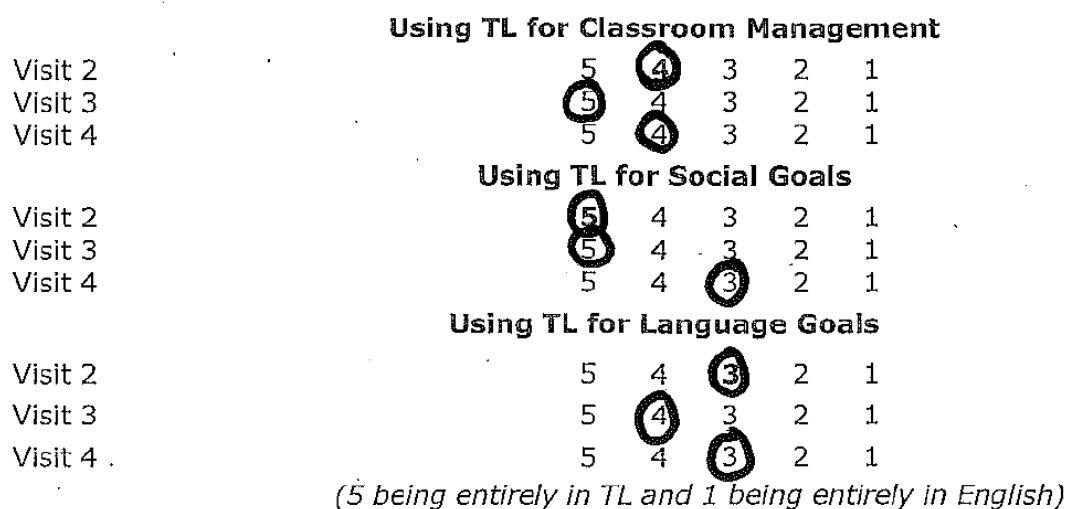


Figure 2. Teacher use of the TL as recorded on EPS document (Visits 2 to 4)

The ISSF rated the criterion of 'teacher use of the target language' as accelerated progress for Visits 2 and 3, and as 'expected progress' for Visit 4 (see Appendix 1).

The researcher conducted her own analysis of the audio-recorded data she had collected during visits to Jane's classroom in order to ascertain teacher use of the *target language* over this time period. She selected lesson excerpts and coded teacher utterances in AS units (Analysis of Speech Units) (Foster et al., 2000). The percentage of these units that were in the TL was calculated. In selecting excerpts, only sections that met Kim and Elder's (2005) criteria, as follows, were chosen:

- interactions between teacher and students during whole class teaching included
- administrative talk at the beginning of the lesson excluded
- sequences involving mechanical TL utterances such as dictations, repetition drills, songs or reading the textbook excluded from consideration

The researcher chose two 5 minute sequences from each lesson; each sequence started with an introduction of a new topic or activity. However for the pre-program first lesson observed, prior to Jane's participation in TPD, and for Visit 1, only 5 minutes met the criteria already mentioned. Table 3 shows percentages of utterances (AS units) in the TL for all visits.

Table 3. Teacher use of Target Language

	Pre-program visit/ 20/3	Visit 1 27/3	Visit 2 15/5	Visit 3 27/7	Visit 4 13/11
First 5 minute excerpt – beginning of lessons					
Percentage of AS units in the TL	14%	34%	90%	97%	76%
Another 5 minute excerpt later in the lesson (visits 2 – 4 only)					
Percentage of AS units in the TL			90%	97%	63%
Total	14%	34%	90%	97%	70%

In the interview the researcher asked Jane the following question:

How important do you think it is for you to try and use French in the classroom as much as possible?

Jane gave similar answers to the pre-TPDL participation interview (1.) and the post-TPDL participation interview (2.).

1. *Well for me all the time cause they're coming to a French lesson.*
2. *Absolutely, yeah, all the time*

However, in the second interview she seems to have changed in her belief that things that she previously thought needed to be said in English, could, in fact, be said in French:

J: It's always been my intention [to use the target language] but then because of restraints of having to explain things I felt, in English, I always refer back to English. But then with TPDL again after the first observation with [ISSF] I thought, right, yes, I'll do that immersion in French.

R: And how did that go?

J: Good, they were great, very responsive . . .

However, Jane indicates that she also thinks that some things need to be said in English:

J: . . . in my last lesson . . . I did use more English . . . just because . . . it's like management, managing the classroom as well as getting them to organise themselves quickly, do it quickly

A little later in the interview Jane identifies increased use of the Target Language as a key goal for the next year, suggesting that she believes it may be possible to use the TL for aspects of classroom management (see underlining added for emphasis in the following extract).

R: So what are you going to do differently next year, if anything?

H: Well I'm going to certainly try and use [French] more, like do as much immersion as I possibly can. I'd do that even with the younger ones with the routines. . . so . . . I'll just say to them in French 'get your pencil cases out, have you got a rubber, have you got a glue stick, use your scissors, cut here'. So all those things are done in French and they understand them, they might not be able to say them or confidently but they know what I'm saying.

Student use of target language

The foci of ISSF Visits 3 and 4 were the students' use of TL in meaningful contexts. As explained previously, the ISSF wrote down all utterances that she heard from two specific students during the lesson. After the lesson the ISSF and the teacher identified together and colour coded those student utterances that were said in the target language. Discussion was aimed at interpreting this evidence and helping the teacher to see what changes might need to be made to increase student use of the TL. An asterisk was placed against student utterances that could have been said in the target language. A short excerpt showing two students' use of language in Jane's class is presented in Table 4.

Table 4. Excerpt of student utterances and coding from ISSF Visit 3

Student one	Student two
*Yeah I think that it should be English	
*Okay come on	
as-tu?	
*you go	
yeah	
un crayon?	merci
*Is it my turn?	un crayon?
un baton de colle?	
	un crayon?
oui	une règle?
merci	des ciseaux?

The researcher used the same audio-recordings made of teacher utterances (given that these were sensitive enough to also pick up learner utterances) to gauge qualitative gains in student use of the TL. She listed all learner utterances that were not just repetitions of language modelled by the teacher. This was either language

that was elicited by the teacher or used independently by students. She then conducted a simple measure of complexity, counting the number of different verb forms that were evidenced in this 'learner language' during each lesson (Yuan & Ellis, 2003). This data, along with information about the different number of formulaic sequences used by learners (as collected by the ISSF and verified by the researcher from her data) is presented in Table 5.

Table 5. Use of Different Verb Forms and of Formulaic Sequences in Student talk

	Pre-programme visit	Visit 1	Visit 2	Visit 3	Visit 4
Total no of different verb forms used/researcher data	1	5	9	13	8
Total no of formulaic expressions used/ISSF data		4	11	12	9

The ISSF noted that students were using single words/ short phrases and full clauses during Visits 2, 3 and 4. There were no examples of multiple clause sentences during any of these lessons.

Provision of opportunities for students to interact in the target language and use it as a tool for meaningful communication

At the end of Visit 1, as has already been explained, the ISSF and Jane completed a 'reflection' which allowed her to evaluate her teaching in relation to Ellis's principles. Question 8 of this 'reflection' was relevant to the above core practice (see Figure 3):

Question 8

Interaction

Do students work in pairs YES/NO In groups YES/NO?
Do they use TL or English to negotiate meaning?

Figure 3. Coding for teacher creation of opportunities for student interaction (Visit 1)

On the EPS document both principles 8 and 2 were written as one key criterion (no. 5 on the EPS document), see below:

These opportunities and communicative tasks allowed the students to use the language as a tool for communication, to initiate interactions (in pairs or groups), find their own words and to negotiate meaning.

The following show the ISSF notes that recorded how Jane's instruction measured against this criterion in Visits 2 to 4:

Visit 2: YES to pair work, YES to using formulaic expression for the memory game, NO to finding their own words

Visit 3: YES – in pair work; YES used language to ask for and give objects; YES- one girl asked for clarification in TL

Visit 4: yes for communication in pairs, no gap

In order to ascertain whether there were increased opportunities for students to interact in the TL the researcher first noted all instances of group/pair work. Table 6 details evidence of any opportunities provided.

Table 6. Opportunities for students to interact in the target language

	Pre-program visit	Visit 1	Visit 2	Visit 3	Visit 4
Opportunity for interaction in TL	no	yes/limited	yes	yes	yes

Table 6 shows that in all four visits there was provision made for learners to interact in groups, something that had not happened in the pre-programme visit.

In the interview the researcher asked Jane the following question:

'Do you get the students sometimes to work in groups/use group work much?'

Once again she gave similar answers to the pre-TPDL participation interview (1.) and the post-TPDL participation interview (2.).

1. *'yes, yeah'*

2. *'yeah I use it a lot and they like that'*

However, later in the interview, she seemed to indicate that she would be using pair work more as she sought to be less teacher centred in her approach. The relevant extract follows:

R: 'what was the most important thing that you feel that you learnt during TPDL?'

J: Not to be teacher controlled, I think within a lesson, they have to work independently on something themselves, you know I can support them. . . that gives me time to go round and listen to them in groups or individually or in pairs. So I think that made me more aware of just taking that step backwards and

setting them up to do something independently without my controlling everything. I think I was probably, well I know, I was quite controlling.

R: Does that mean that you're using more groups now as a result of TPDL, more group and pair work?

J: Certainly more pair work, yeah definitely more pair work. . . . So between them they would be coming up with the language more. But certainly pair work, I think, yeah, cos they need to be challenged and extended and I think that's a good way of doing it

Looking then at the impact of the programme on Jane's classroom practice, we can see that there was a dramatic and immediate increase in the *teacher's use of target language*. By Visit 2 Jane was using the target language in 90% of utterances (coded as AS units) and at Visit 3 this had increased further to 97%. This was a huge increase from the 24% average of the pre-programme visit and Visit 1. However, in Visit 4, the use of TL was less evident, at a 70% average. As documented earlier, Jane referred, in the interview, to the fact that she had used less TL in the last visit and seemed to attribute this to finding that English was necessary to get the students to do what she intended quickly. It seems that she had, perhaps, lost the conviction she talked about after the first observation that she could do 'immersion in French'. Jane does, however, identify use of the target language as a key goal going forward and indicates that she will use French for 'routines' in the classroom.

There were also gains in *student use of the target language*. The researcher heard one verb form used by learners during the pre-programme visit, this had increased to 13 different verb forms in Visit 3. There was a slight decline in both measures of student language use in Visit 4 (it is interesting to see how decreased student use of TL accompanied a reduction in teacher use of the same). In interpreting this data it is important to remember that Jane taught these students for one 50 minute lesson a week only. We also do not have data that would tell us the qualitative gains in language use that students whose teachers were not involved in the TPDL programme might have made over the same period with the same amount of input.

Jane also provided *opportunities for students to interact in the TL* in all four ISSF visits, although this was 'limited' in Visit 1 and there was no occasion for this in the pre-programme visit, suggesting that Jane did make this a greater priority in response to programme demands. Jane also described herself as a 'controlling teacher' and expressed a wish to give more independence to students in terms of opportunities for using language to interact and communicate meaning.

In summary then, this data from Jane's classroom constitutes some evidence for the positive impact of the evaluation process on teaching practice, thus demonstrating

positive backwash and some evidence of consequential validity. For example, Jane claimed, in the pre-programme interview, that it was important to use target language in the classroom 'all the time'. It is to be presumed, however, that confronting her with evidence from Visit 1 that, in her classroom, this was a long way from being the reality, may have created the 'cognitive dissonance' (Dangel & Guyton, 2004; Kagan, 1992) needed as an incentive for her to change her practice. In any event it was, she claimed, the conversation with the ISSF that acted as the catalyst for her to try using 'immersion'.

Ongoing research is needed to ascertain to what extent changes in behaviour are sustained, to investigate impact of the programme in a wide variety of teaching contexts, and, in particular to investigate to what extent student learning is impacted. Another consideration is the extent to which the core practices that underpin the TPDL program are of value in all teaching contexts.

Limitations

This paper has described the implementation of a small part of the evaluation process for one of the teachers that took part in TPDL in 2014. It is thus limited in the insight that it gives. For example, it does not provide information about how the discussion and reflection process took place in the dialogic interaction between the ISSF and the teacher participant. From such a restricted case study it is also impossible to generalise to the experience of other TPDL participants.

As described above, there is some evidence for the impact of the programme on teaching practice but more research is indicated. The low 'buy-in' from TPDL participants in the research project referred to in the paper highlights one of the potential difficulties of this type of research, that teachers may tend to see such research initiatives as an evaluation of their own teaching practice, rather than as an attempt to estimate the value of the evaluation tools used in the programme in order to make them more fit for purpose and thus to contribute to their ongoing validation.

Conclusion

This paper has described and documented the implementation of two evaluation tools designed to promote change in teacher practice as part of an in-service language teacher professional development programme. It has described how these evaluation tools were informed by theoretical understanding about what makes for effective language learning and how this understanding was used to identify core practices that, it was hypothesized, would impact positively on student learning

(Kubaniyiova & Feryok, 2015). It documents the emphasis that was placed on encouraging teachers in the programme to analyse and reflect on their own practice and the increasing responsibility that they undertook for this over the course of the programme (Scarino, 2014). It provides some evidence of change in teacher practice for the three key criteria outlined in the Progress Standards document. While this paper has therefore documented a range of evidence for the validity of these tools, further research is indicated that could collect additional and ongoing evidence for their validity and, as part of this process, continue to investigate to what extent they may lead to a positive impact on teacher practice and on student learning. Future research would also do well to consider other aspects of the evaluation process, such as the 'learning discussion' that takes place between the ISSF and the teacher.

References

- Ball, D., & Forzani, F. (2009). The work of teaching and the challenge for teacher education. *Journal of Teacher Education*, 60, 497-511.
- Byram, M. (1997). *Teaching and Assessing Intercultural Competence in Language Teaching*. Clevedon: Multilingual Matters.
- Dangel, J. R., and Guyton, E. (2004). An emerging picture of constructivist teacher education. *The Constructivist*, 15, 1-35.
- Ellis, R. (2005a). *Instructed second language acquisition: A literature review*. Wellington: Research Division, Ministry of Education.
- Ellis, R. (2005b). Principles of instructed language learning. *System*, 33(2), 209-224.
- Erlam, R. (2008). What do you researchers know about language teaching? Bridging the gap between SLA research and language pedagogy. *Innovation in Language Learning and Teaching*, 2, 3, 253-267.
- Erlam, R., & Sakui, K. (2006). *Instructed second language acquisition case studies*. Wellington: Ministry of Education.
- Foster, P., Tonkyn, A. and Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21, 354-375.
- Freeman, D., McBee Orzulak, M., & Morrissey, G. (2009). Assessment in second language teacher education. In A. Burns & Jack C. Richards (Eds.), *The Cambridge Guide to Second Language Teacher Education* (77-90). New York, NY: Cambridge University Press.
- Gibbs, C.J., & Holt, R.F. (2003). *Teaching of international languages in New Zealand schools in years 7 and 8: An evaluation study*. Wellington: Ministry of Education.
- Hiebert, J., & Morris, A. (2012). Teaching, rather than teachers, as a path toward improving classroom instruction. *Journal of Teacher Education*, 64, 92-102.

- Hughes, A. (2003). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Insley, S., & Thomson, W. (2008). Supporting teachers to increase the quality and quantity of student target language use: Description and analysis of the 'Teacher Professional Development in Languages' (TPDL) programme. *The New Zealand Language Teacher*, 34, 21-30.
- Kagan, D. (1992). Professional growth among preservice and beginning teachers. *Review of Educational Research*, 62, 129-169.
- Kim, S. H., & Elder, C. (2005). Language choices and pedagogic functions in the foreign language classroom: A cross-linguistic functional analysis of teacher talk. *Language Teaching Research*, 9, 355-380.
- Kubanyiova, M., & Feryok, A. (2015). Language teacher cognition in applied linguistics research: Revisiting the territory, redrawing the boundaries, reclaiming the relevance. *Modern Language Journal*, 99(3), 435-449.
- Mathew, R., & Poehner, M. E. (2014). Monitoring progress in the classroom. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp. 631-645). New Jersey: John Wiley & Sons.
- McDonald, B., & Boud, D. (2003). The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education*, 10(2), 209-220.
- Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (pp. 13-103). New York, NY: Macmillan.
- Ministry of Education. (2007). *The New Zealand Curriculum*. Wellington, New Zealand: Learning Media.
- Newton, J., Yates, E., Shearn, S., & Nowitzki, W. (2010). Intercultural communicative language teaching: implications for effective teaching and learning. (pp. 1-90). Report to the Ministry of Education. <http://www.educationcounts.govt.nz/publications/curriculum/76637/introduction>
- Rivers, J. (2010). *An introduction to the concept of intercultural communicative language learning and teaching: A summary for language teachers*. New Zealand: Ministry of Education.
- Ross, J. A. & Bruce, C.D. (2007). Teacher self-assessment: A mechanism for facilitating professional growth. *Teaching and Teacher Education*, 23(2), 146-159.
- Ross, J. A., Hogaboam-Gray, A., & Rolheiser, C. (2002). Student self-evaluation in grade 5-6 mathematics: Effects on problem solving achievement. *Educational Assessment*, 8(1), 43-58.
- Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (1999). Effects of self-evaluation training on narrative writing. *Assessing Writing*, 6(1), 107-132.

- Scarino, A. (2014). Learning as reciprocal, interpretive meaning-making: A view from collaborative research into the professional learning of teachers of languages. *The Modern Language Journal*, 98, 386-401.
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Harlow: Pearson Education.
- TPDL. (2014). *Teacher professional development languages participant handbook*. Auckland, New Zealand: The University of Auckland.
- Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. New York, NY: Palgrave Macmillan.
- Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, 24, 1-27.
- Zeichner, K. (2012). The turn once again toward practice-based teacher education. *Journal of Teacher Education*, 64, 376-382.

Appendix 1

In-School Support Measuring Progress/Progress Standards

Visit 1 provided the baseline picture from which to measure your progress during the TPDL programme. For subsequent visits, there are three expected standards of progress for you to meet.

4. Sets the progress standard for teacher use of the TL.
5. Sets the progress standard for student use of the TL.
6. Sets the progress standard for the provision of opportunities for student interaction in the TL.

Progress Standards for Visits 2, 3 and 4

Visit 2

Expected progress:

1. Teachers are using TL for social goals, classroom management and instruction.
2. Students show evidence of using the TL in meaningful classroom contexts.
3. Students work in pairs or groups to complete language tasks.

Accelerated progress:

1. Teachers are making significant use of the TL for social goals, classroom management and instruction.
2. Students show evidence of using the TL in meaningful classroom contexts and their TL utterances are more than just words and short phrases.
3. There is evidence of students using strategies to cope in the TL.

Visit 3

Expected progress:

1. Teachers are making significant use of the TL for social goals, classroom management and instruction.
2. Student utterances are quantitatively and qualitatively more sophisticated than was evident in visit 2.
3. Teachers are providing students with opportunities to use the TL as a tool for communication.

Accelerated progress:

1. Teachers are using the TL for most of their talk with students in a way which is comprehensible and motivating. Teacher talk does not dominate the lesson.
2. Student utterances are quantitatively and qualitatively more sophisticated than was evident in visit 2 and student use of the TL includes multi-clause conversation.
3. Teachers are providing students with opportunities to use the TL as a tool for communication, with successful outcomes.

Visit 4

Expected progress:

1. Teachers are making significant use of the TL for social goals, classroom management and instruction. Teacher talk does not dominate the lesson.
2. Student utterances are quantitatively and qualitatively more sophisticated than was evident in visit 3.
3. Students are using the TL to negotiate meaning with each other.

Accelerated progress:

1. Teachers are using the TL for most of their talk with students in a way which is comprehensible and motivating. Teacher talk does not dominate the lesson.
2. Student use of the TL is predominant and displays a rich repertoire of formulaic expressions and multi-clause conversation in the TL.
3. Students are taking ownership of their interactions in the TL.