

## INTRODUCTION

## EINLEITUNG

## ● Laura Loder Büchel | Zürich

Laura Loder Büchel is a lecturer at Zurich University of Teacher Education and a member of Babylonia's editorial staff.



I am constantly struggling to make meaning of how my own children are being assessed in Swiss primary and secondary schools and I am sure I am any teacher's nightmare when I come with a question about such practices. How can an American who only learned her first foreign language at the age of 20 even possibly fathom the depth and breadth of foreign language assessment in the multilingual Swiss context? There are many things going right when my Swiss-German speaking children come home motivated to learn, to study, to use their French when we visit Geneva, to scream a French poem at me from another room, to add a good piece to their portfolio, to rewrite something for their portfolio. So how can I possibly even consider the following dark thoughts (see Table 1 on the next page) when so much is right? Luckily, I am not the only one having these thoughts; the mass of literature found will attest that is a hot topic with mismatches between what teachers say they are doing in terms of assessment and what they are actually doing (Pat-El, Tillema, Segers & Vedder, 2015). There

Täglich stelle ich mir tausend Fragen darüber, wie meine Kinder in der Primar- und Sekundarstufe in den Fremdsprachen Englisch und Französisch - und auch in der Schulsprache Deutsch - beurteilt werden. Wie kann eine Amerikanerin, die ihre erste Fremdsprache erst im Alter von 20 gelernt hat, die vielfältigen Dimensionen von Beurteilungen im mehrsprachigen Schweizer Kontext nachvollziehen? Vieles läuft gut, wenn meine deutschsprachigen Kinder motiviert von der Schule nach Hause kommen, um die „Ufzgi“ ohne Jammern zu machen, oder wenn sie versuchen, bei einem Besuch in Genf ihr Französisch zu sprechen, wenn sie lustige Gedichte auf Französisch laut vorsagen, wenn sie stolz etwas Neues im Portfolio eintragen und ihre Texte mit Freude bearbeiten. Also: Wie kann es sein, dass ich immer wieder finstere Gedanken habe (vgl. *Table 1* auf der nächsten Seite), auch wenn so vieles richtig läuft? Glücklicherweise bin ich überhaupt nicht alleine mit meinen finsternen Gedanken: Eine umfassende Literatur bestätigt, dass Beurteilen und Bewerten ein echtes Spannungsfeld ist, mit Diskrepanzen

Table 7

Dark Thoughts....		Finstere Gedanken	
Mommy me	Lecturer me	die Mutter in mir	die Dozentin in mir
Why were points taken off on my daughter's English test when she spelled a German word incorrectly?	Why are children even being given vocabulary translation tests when most kids will certainly not become translators/interpreters? What is actually being measured through these tests?	Warum hat meine Tochter weniger Punkte für einen Englischtest erhalten nur wegen ihren deutschen Rechtschreibfehlern?	Warum machen Lernende Übertzungstests, obschon nur die wenigsten Übersetzer werden? Was sind die Konstrukte hinter Übertzungstests?
Why do they have to learn old and even incorrect words for things like professions (scheduler? warehouseman?)? Why is the writing test on the use of the present perfect and not on some specific text type (like an informal letter)? What else can they do in the foreign language?	Are schools actually measuring the skills that learners will need in the future? How can we better teach functional literacy skills in English that might even transfer to general skills (writing letters, giving instructions, etc...)	Warum lernen meine Kinder veraltete Wörter wie "warehouseman"? Warum konzentrieren sich Schreibtests auf Grammatik (wie "Present Perfect") und nicht auf Textsorten (wie Einladungen) oder Satzstrukturen? Was können sie sonst noch in der Fremdsprache?	Beurteilen Schulen die nötigen Fertigkeiten, welche Lernende in Zukunft und im täglichen Leben tatsächlich einsetzen werden? Wie können schriftliche Kompetenzen im Englischunterricht besser unterrichtet werden, damit sie vielleicht sogar übertragbar werden (Briefe schreiben, Anweisungen geben)?
Why did I receive a sneak peek at my daughter's report card grades from the school's data collection tool and there was no separation of subskills (reading, writing, speaking, listening)?	How can one actually separate the four skills for testing purposes when even a gap fill measures reading and writing? How is interaction measured? How can teachers have good systems for recording development in each skill?	Kurz vor der Zeugnisabgabe gewährte mir die Lehrperson Einblick ins Notenraster meiner Tochter. Warum habt sie nur die Durchschnitte aller Prüfungen genommen und nicht die vier Fertigkeiten berücksichtigt?	Wie könnten die vier Fertigkeiten (Lesen, Schreiben, Sprechen, Zuhören) isoliert bewertet werden, zumal beispielsweise ein Lückentext sowohl Lesewie Schreibkompetenz verlangt? Wie lässt sich „Interaktion“ bewerten? Wie könnten Lehrende die sprachliche Entwicklung einzelner Fertigkeiten dokumentieren?
When I asked the teacher what evidence s/he had collected about my daughters' performances, I got "here are her tests" and no other indications of having met any standards.	Where are the models for data collection (such as a grade being 20% average of 4 tests, 40% average of 8 performance assessments such as essays, 20% average of 20 pop quizzes...) and why are they not being used for grades?	Frage ich die Lehrerin nach Beispielen für die Leistungen meiner Tochter, erhalte ich als Antwort "Hier sind die Prüfungen" und keine anderen Hinweise darauf, ob sie Lernziele erreicht hat.	Was sind nützliche Modelle für evidenzbasierte Zeugnisnoten (z.B. dass 4 Prüfungen durchschnittlich 20% zählen, dass Leistungen wie Essays 40% ausmachen usw.) und warum werden sie nicht für die Beurteilung genutzt?
My daughter came home with a "True/False/Not in the text" test that even I got wrong.	Why are test-designers not aware of "True/False/Justify" and better test item design?	Meine Tochter zeigt mir eine "Richtig/Falsch/Keine Aussage" Englischprüfung, die selbst ich als Muttersprachlerin nicht richtig beantworten könnte.	Warum haben Lehrmittelentwickler oft beschränkte Repertoires von Test Items, die verwendet werden können? Wäre die Formel „Richtig/Falsch/Rechtfertigen“ nicht die bessere Alternative zu „Richtig/Falsch/Keine Aussage“?
Where are the stories, the arts, the intercultural skills, the strategic competencies that my daughters might have to learn about or to use? I mean, both are pretty good at talking with their hands!	How can we measure such skills when there is often no right or wrong? How can the ability to "negotiate meaning" and the development of strategies be assessed and reported?	Wo sind die Geschichten, die Kunst, die interkulturellen Fertigkeiten, die strategischen Kompetenzen, die meine Töchter lernen und einsetzen könnten? Schliesslich können sich beide gut mit Gesten ausdrücken!	Wie können wir Fertigkeiten evaluieren, für die es kein "gut" oder "unzureichend" gibt? Wie kann das „Aushandeln von Bedeutungen“ und die Entwicklung von Strategien beurteilt und bewertet werden?
Despite all the "hoopla" around "Kompetenzorientierung", has standards-based grading reached most classrooms – and should it?		Ist standardbezogene Bewertung an Schweizer Schulen schon angekommen, trotz des Tamtams rund um "Kompetenzorientierung"?	

The mass of literature found will attest that is a hot topic with mismatches between what teachers say they are doing in terms of assessment and what they are actually doing.

is a clear need for improvement and development and assessment-literacy is an often-overlooked aspect of teacher training as well as continuing practice (Popham, 2011). DeLuca, Chavez, Bellara & Cao (2013) found four explicit constructs that aid in teacher assessment literacy development:

- › perspective-building conversations,
- › praxis activities,
- › modeling, and
- › critical reflection and planning for learning.

These constructs are alive and kicking in this issue of *Babylonia* as we take a look at perspectives and models from the US (see *Howard on Class Dojo*), from England (see *Jones on Assessment for Learning*), from Sweden (see *Snoder on the new Swedish curriculum*) from Germany (see *Rymarczyk on the importance of writing*), from Holland (see *Fasoglio et al.*) and return back home to Swiss turf (see *Stotz on the Stellwerk and Multicheck tests*, *Bartholemy, Bolomey & Schweitzer on the new German language test in the canton of Vaud*, *Hofer on the importance of language at Manpower*) whilst travelling through the ever-changing turbulent landscape of grading and assessment policies and practices. This issue of *Babylonia* can be supportive in light of recent suggestions for new school report cards (as *Amsler* mentions) and a turn towards standards-based grading (see *Kuendig or Büchel or Meili for some concrete ideas*) and provides myriad ideas for rethinking and refining practices and promoting discussion. It can help classroom teachers to examine how to measure intercultural skills (see *Eberhardt*), test listening and reading comprehension (see *de Pietro & Roth on testing gist* or *Barras, Karges & Lenz on testing reading* or *Rothenberger on literature*), write valid tests (see *Studer on two useful test item types* or *Torresan on testing Italian language skills as well as Karges who provides a checklist for teachers when writing tests*) and it provides numerous praxis activities. It can encourage instructors to critically think about what messages are being passed on to future and in-service teachers.

So, enjoy the roller-coaster ride through Tema 1 (*assessing receptive skills*), Tema 2 (*testing in general*) and Tema 3 (*practical ideas*) and feel free to use the Reader's Guide questions to promote discussion.

zwischen dem, was Lehrkräfte zu machen scheinen, und dem, was tatsächlich gemacht wird (e.g. Pat-El, Tillema, Segers & Vedder, 2015). Es gibt Spielraum für Verbesserungen und Entwicklung bei den Beurteilungskompetenzen von Lehrpersonen, was leider oft in der Aus- und Weiterbildung übersehen wird (vgl. Popham, 2011). DeLuca, Chavez, Bellara & Cao (2013) definierten vier explizite Konstrukte, die bei der Entwicklung von Beurteilungskompetenzen helfen können:

1. Austausch zwischen Lehrpersonen über Beurteilungspraxis (*perspective-building conversations*);
2. Praktische Aufgaben (*praxis activities*);
3. Modellierung (*modeling*); und
4. Kritische Reflexion und Planung von Unterrichtseinheiten (*critical reflection and planning for learning*).

Diese vier Modelle sind in dieser Nummer gut vertreten, indem wir auf Perspektiven und Modelle aus den USA (*Howard zu Class Dojo*), aus England (*Jones zu Assessment für Learning*), Schweden (*Snoder zum neuen schwedischen Lehrplan*), Deutschland (*Rymarczyk zum Stellenwert des Schreibens*) und Holland (*Fasoglio et al.*) eingehen, um dann auf dieser Reise durch die sich ständig veränderten Landschaften der Beurteilungs- und Prüfungspraxis in die Schweiz zurückzukehren: *Stotz* berichtet über die Stellwerk- und Multicheck-Tests; *Bartholemy, Bolomey & Schweitzer* über die neue Deutschprüfung im Kanton Waadt und *Hofer* über Sprachen bei *Manpower*. In dieser Ausgabe von *Babylonia* finden sich Empfehlungen für neue Schulzeugnisnoten (*Amsler*), Vorschläge zu standardbasierter Beurteilung (vgl. *Kuendig oder Büchel oder Meili*) und eine Vielzahl weiterer Ideen, die dazu anregen, die Praxis zu überdenken und zu verfeinern. Klassenlehrpersonen werden von einer Vielfalt an Praxisideen profitieren, wie zum Beispiel zur Bewertung von interkulturellen Fertigkeiten (vgl. *Eberhardt*), zur Beurteilung von Lese- und Hörverständnis (siehe *de Pietro & Roth oder Rothenberger*) und zur Testvalidierung (vgl. *Studer, der zwei neue Testformate beschreibt, oder Torresan, zu Sprachkompetenztests für Italienisch und Karges, die eine Checkliste zur Testerstellung für Lehrende vorstellt.*). Mit dieser Nummer werden Dozierende ermutigt, kritisch darüber nachzudenken, welche Botschaften

sie zukünftigen und bereits im Beruf tätigen Lehrenden vermitteln.

Geniessen Sie die Achterbahnreise durch Tema 1 („Beurteilung von rezeptiven Kompetenzen“), Tema 2 (Sprachtests allgemein) und Tema 3 („Praxis“). Zur Förderung der Diskussion dient bestimmt unser „Reader’s Guide“.

### Reader’s Guide

*You can use this issue of Babylonia for discussing the topic with fellow teachers and students. Here are some questions to discuss before or after reading this issue of Babylonia!*

- > What does testing in a competence-oriented way mean? How can tests be written to reflect WHAT learners should be able to do in the language?
- > What systems are you (teachers) using to have an evidence-based approach to assessing both receptive and productive skills? How do you gather evidence? Did you pick up on any new tools or ideas that you’d like to try out?
- > Teachers in some cantons are asked to report on their learners’ reading, writing, speaking and listening skills in isolation on school report cards. How, thus, can they measure “interaction” and also separate the skills for these reports?
- > What constructs can you find mentioned for: listening comprehension, intercultural skills, literature studies, a good test? When you are focusing on a specific skill in the classroom, how do you make learners take ownership for what is being measured?
- > One could argue that testing in schools is not even necessary, and that performance assessments (presentations, essays, dialogues), good note-taking habits and smaller quizzes would be enough – do you agree?
- > What makes a good reading, writing, speaking or listening test?
- > What do vocabulary-translation tests measure? From this issue of Babylonia, can you make a list of other techniques that might provide a more in-depth picture of learner knowledge and skill?

### References

- DeLuca, C., Chavez, T., Bellara, A. & Cao, C.** (2013). Pedagogies for preservice assessment education: Supporting teacher candidates’ assessment literacy development. *The Teacher Educator*, 48: 2, 128-142.
- Pat-El, R., Tillema, H., Segers, M. & Vedder, P.** (2015). Multilevel predictors of differing perceptions of Assessment for Learning practices between teachers and students. *Assessment in Education: Principles, Policy & Practice*. 22, 2.
- Popham, W. J.** (2011). Assessment literacy overlooked: A teacher educator’s confession. *The Teacher Educator*, 46(4), 265-273.