

Forschungsarbeiten zur Validierung von rezeptiven Testaufgaben

Peter LENZ

Kompetenzzentrum für Mehrsprachigkeit
Universität Freiburg/Schweiz

Ringvorlesung, 18.12.2017

Content

Language test validation

Definitions

Methods

Research in test validation

The language of questions and answers (*Task Lab*)

Outlook (*Innovative Forms of Assessment*)

Language test validation

DEFINITIONS

Validity for measurement

[An assessment, test, examination, task ... is] a set of operations in which behaviour is elicited, evaluated and interpreted; often, the outcome [...] will be summarized as a score, result or profile report, which is intended to characterize the individual (or higher-level entity, e.g. school) in terms of the attribute being measured.

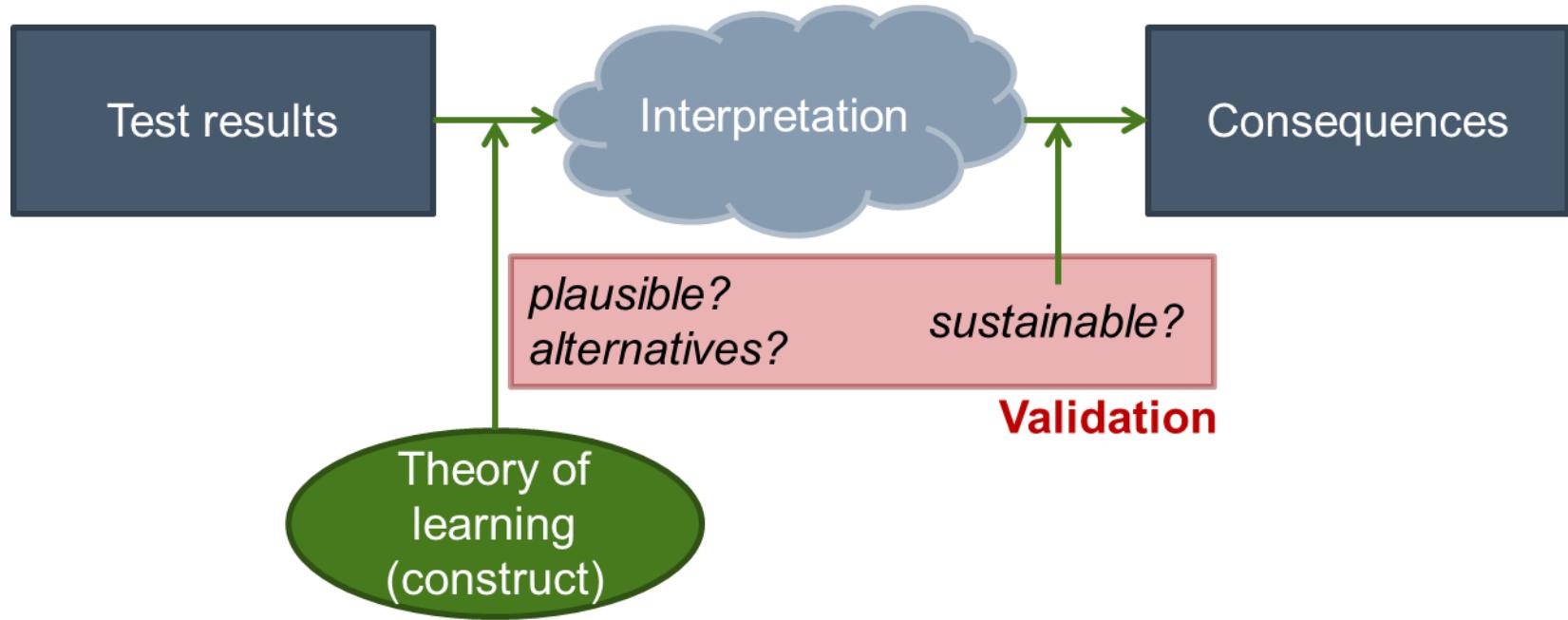
Newton & Shaw 2014:10

Validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment.

Messick 1989:13

Validation

Kane 2006, 2012; McNamara 2010; Messick 1989, 1990



Language test validation

METHODS

Pre-operational testing

(...) tests that evaluate language users' abilities need to be evaluated themselves.
(Chapelle & Voss 2014:1081)

*Pre-operational testing is the thoughtful, deliberate, and systematic process of **collecting and analyzing evidence to support the validity of an assessment prior to that assessment's operational use***
(Kenyon & MacGregor, 2012:295).

→ “an overlooked topic” and “the underappreciated step-child of the test development process” (Kenyon & MacGregor 2012:305)

Objectives of pre-operational testing

- To investigate the functioning of items, tasks or the test as a whole.
- To discover how examinees respond to tasks (test-taking processes, perceptions of tasks, performance).
- To modify the test based on the results.

(Kenyon & MacGregor 2012)

Pre-operational testing methods

- Qualitative methods (*pilot studies*)
 - Formal cognitive laboratories
 - Observation
 - Think-aloud protocols
 - Interviews
 - ...
 - Quantitative methods (*field testing*)
 - Psychometrical analysis of item and test functioning
- (Kenyon & MacGregor 2012)

Qualitative methods in language test development

- Increase since the 1980s (complementary to the well-known quantitative methods)
- Possible explanations for this increase:
 - McNamara, Hill & May 2002: Cross-fertilisation of research in language testing and other areas of applied linguistics
 - Banerjee & Luoma 1997, Bachmann 2000: Widening conceptualisation of validity in the assessment literature in general and in language testing in particular (Messick 1988, 1989)

Verbal protocol analysis (VPA)

- A *protocol* is a “*verbal report*”, or set of utterances, collected under special conditions and constituting a body of qualitative data for detailed analysis (Lazaraton & Taylor 2007:122).
- The assumption underlying introspection, in general, is that **it is possible to tap into and document a learner's internal processes** in much the same way as one can observe external real-world events. This is predicated on an additional assumption, namely that **humans have access at some level to their internal thought processes and can verbalize those processes** (Gass & Mackey 2017:1).
- If used appropriately: a valid, reliable and useful technique (Green 1998, Bowles 2010, Gass & Mackey 2017).

Variations on the verbal report procedure

Form of Report

TALK ALOUD

THINK ALOUD

Temporal Variations

Concurrent

Retrospective

Concurrent

Retrospective

Procedural Variations

Mediated

Non-Mediated

Mediated

Non-Mediated

Mediated

Non-Mediated

Mediated

Non-Mediated

Green 1998: 5

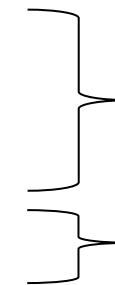
VPA in language testing

- Used as a supplement to the quantitative data analysis during the test validation process.
- Used to investigate the strategies involved in **test-taking** and **rating**.
- In focus: **validity of the instrument** and **reliability of judgements**.

Research on test-taking strategies (e.g. Cohen 2011, 2014)

(...) strategies that respondents are drawing upon as they complete language tests - that is, the **consciously-selected processes** that the respondents use for dealing with both the **language issues** and the **item response demands** in the test-taking tasks at hand. (Cohen 2011:305)

1. Language learner strategies
2. Test-management strategies
3. Test-wiseness strategies



Construct-relevant variance

Construct-irrelevant variance

[The best way to guard against the use of test-wiseness strategies is to first] *collect data that accurately reflect what respondents do to respond to items and tasks, and then to construct items and tasks that require the use of language skills in order to produce correct answers* (Cohen 2014:901).

The potential of VPA in language testing

Verbal protocol analysis (...) provides us with a means for both testing and developing theories. In the context of language testing, it has the potential to elucidate the abilities that need to be measured, and also to provide a means for identifying relevant test methods and selecting appropriate test content.

Protocol analysis should not be seen merely as a method to be used as part of construct validation, but should also be viewed as a means by which we may enrich our understanding of the very nature of the construct(s) under investigation (Green 1998:120).

Disadvantages of VPA

- **Time consuming**
- Procedural difficulties: it is easy to unintentionally disrupt or alter participants' behaviour
- Individual differences among the participants

Data analysis

- Transcription of all interviews
- Analysis/coding using MAXQDA
- Content analysis (cf. Mayring)

Quantitative methods: Relevance

- *Quantitative item and test analysis* has a long-standing tradition, particularly in Anglo-Saxon professional language testing (and, of course, psychological testing, educational measurement, etc.).
- In *applied linguistics*, actual development and validation of measurement instruments is often ignored or neglected, but this is changing (cf. Purpura, Brown & Schoonen (2015) in *LL*).
- Quantitative (psychometric) approaches to item and test validation can
 - make the properties of measurement instruments and the properties of person measures based on them known,
 - improve measurement instruments,
 - add to the validity evidence from qualitative sources.

Construct validity from a quantitative perspective

- Measurement instrument is unidimensional or controlled multidimensional
- Instrument has specific objectivity (measures persons with 'more' or 'less' of the construct equally along the scale)
- Test-takers are classified similarly when tested on a recognized instrument (criterion) measuring the same construct.

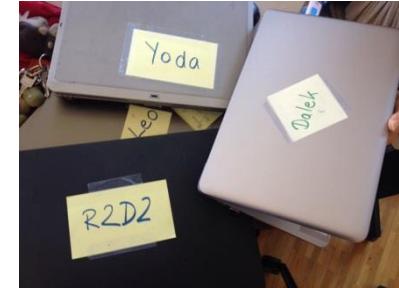
Questions for analysis:

- Are the items that are assumed to be equal – equal?
- Do all of the items and item groups in the instrument 'fit', i.e. function according to the model used across all items?
- Do the items perform equally for relevant groups of users? (DIF)

THE TASK LAB PROJECT

Task Lab

- **Objectives**
 - Understand computer-based reading assessment tasks better
 - Investigate controversial task features
 - item types
 - language of rubrics and responses
 - Gain insight into the interaction between test-taker characteristics, task features and success on items
- => Inform test development for large-scale assessments
- => Inform assessment scale interpretation
- **Target group:** pupils age 12, German = language of schooling; French = first foreign language, 4 years of instruction, 2 or 3 lessons per week)



Instruments

Reading tasks:

- 12 tasks in 4 versions (3 items each):
 - MCQ + German
 - MCQ + French
 - SAQ + German
 - SAQ + French
- 6 tasks in 2 test versions (3 items each):
 - MTCH + German
 - MTCH + French

Component tests:

- Digit span tasks (forward/backward)
- Sight word recognition
- Pronunciation of non-words
- C-Tests
- Segmentation tasks
- Yes/no vocabulary test

Questionnaires

Instruments

Component tests (I)

Was kann für dich gut in Französisch?
 Ich kann in Französisch leicht gut sprechen, viel leichter als schwierig.
 Ich kann in Französisch leicht schreiben.
 Ich kann in Französisch leicht lesen.
 Ich kann in Französisch leicht schreiben und leichter als schwer schreiben.
 Ich kann in Französisch leichter als schwer schreiben.

Deine Familie und du
 In welchem Jahr bist du geboren? 20
 In welchem Land bist du geboren? DE
 Wo wurden deine Eltern geboren? DE
 Deine Mutter: DE
 Dein Vater: DE
 Was ist deine erste Sprache?
 Das ist die Sprache, die ich als Kind gesprochen habe.
 Wenn du gleichzeitig zwei oder mehr Sprachen gesprochen hast, kommt die zweite Sprache automatisch.
 Wenn du gleichzeitig zwei oder mehr Sprachen gesprochen hast, kommt die zweite Sprache automatisch.

Welche Schule hast du bis jetzt besucht?
 alle 7 Jahre
 1. Jahr Kindergarten
 2. Jahr Kindergarten
 1. Klasse Primarschule
 2. Klasse Primarschule
 3. Klasse Primarschule
 4. Klasse Primarschule
 5. Klasse Primarschule



Student Questionnaire

- Gender
- Language background
- Motivation (enjoyment)
- Motivation (ought)

Backward Digit Span Task

Working memory/ processing

Sight-word recognition

Word decoding (gestalt)

Instruments

Component tests (II)



Yes-No Task
Vocabulary breadth (receptive)

Text segmentation
Morpho-syntax & integrative
measure

C-Test
Integrative measure / written
text reconstruction

Instrument development and data collection

- Pre-piloting
 - Retrospective interviews/Stimulated recall for reading tasks (34 students)
- Piloting (field study)
 - Piloting of all instruments (97 students)
- Main data collection
 - 35 classes, ≈ 600 students

Task Lab

THE QUALITATIVE STUDY AND SOME RESULTS

Qualitative study: retrospective interviews

- One student – one researcher
- 45 minutes per student
- Interviews immediately after each item
- Stimulus for interview: reading or listening passage(s)
 - + item + student's answer
 - Questions regarding the reading process
 - Questions regarding usability (screen design etc.)
 - Questions regarding personal preferences, further challenges
- Protocol: audio-recording and written notes

Example: Task 4 / item 1

T04_ls

Un chat entre Timo et Martin

Martin : Salut Timo! Ça va ? Toujours dans la grande capitale japonaise ? A bientôt, Martin.

Timo : Salut ! Oui, on est toujours là, c'est super !!! C'est vraiment incroyable : plus de 13 millions de personnes habitent dans cette ville ! J'adore ça ! Hier on est allé sur la tour « Tokyo Skytree ». Là, on a une magnifique vue sur toute la ville. C'était trop cool !

On mange souvent dans les restaurants : les sushi, « l'udon » (une soupe), et les glaces sont trop bons. Et il y a des automates avec des boissons ou même des « ramen » (des pâtes) partout. Et puis, aujourd'hui, à midi, on a mangé du « takoyaki » dans un restaurant. Ça n'avait pas l'air bon, mais quand j'ai essayé, j'ai changé d'avis : c'était délicieux ! Maintenant, c'est mon plat préféré !

On reste encore deux jours à Tokyo et après on va à Nara où il y a une statue en bronze gigantesque. Après Nara, on va à Okinawa. Là-bas, on a beaucoup de trucs à visiter. Moi, je veux aller au musée du karaté ! Je t'enverrai une photo.

On rentre en Suisse dans 10 jours. A bientôt ! Timo

Links liest du einen Chat zwischen Timo und Martin.
Dazu stellen wir dir drei Fragen.

1. Frage:
Was ist das Thema von Timos Nachricht?

- Ein Dokumentarfilm über Japan
- Seine Ferien in Japan.
- Sein Abend in einem japanischen Restaurant.

Weiter

Fürst Frage, dann
strategisch im
Text gesucht

es geht um japanische
Restaurants
, nach entsprechende
Stichwörter gesucht



Example: Task 4 / Item 2

Un chat entre Timo et Martin

Martin : Salut Timo! Ça va ? Toujours dans la grande capitale japonaise ? A bientôt, Martin.

Timo : Salut ! Oui, on est toujours là, c'est super !!! C'est vraiment incroyable : plus de 13 millions de personnes habitent dans cette ville ! J'adore ça ! Hier on est allé sur la tour « Tokyo Skytree ». Là, on a une magnifique vue sur toute la ville. C'était trop cool !

On mange souvent dans les restaurants : les sushi, « l'udon » (une soupe), et les glaces sont trop bons. Et il y a des automates avec des boissons ou même des « ramen » (des pâtes) partout. Et puis, aujourd'hui, à midi, on a mangé du « takoyaki » dans un restaurant. Ça n'avait pas l'air bon, mais quand j'ai essayé, j'ai changé d'avis : c'était délicieux ! Maintenant, c'est mon plat préféré !

On reste encore deux jours à Tokyo et après on va à Nara où il y a une statue en bronze gigantesque. Après Nara, on va à Okinawa. Là-bas, on a beaucoup de trucs à visiter. Moi, je veux aller au musée du karaté ! Je t'enverrai une photo.
On rentre en Suisse dans 10 jours. A bientôt ! Timo

Links liest du einen Chat zwischen Timo und Martin. Dazu stellen wir dir drei Fragen.

2. Frage:
Welches Essen hat Timo am besten geschmeckt?

Takoyaki
 Sushi
 Glace

[Wieder](#)



Example: Task 4 / Item 3

Un chat entre Timo et Martin

Martin : Salut Timo! Ça va ? Toujours dans la grande capitale japonaise ? A bientôt, Martin.

Timo : Salut ! Oui, on est toujours là, c'est super !!! C'est vraiment incroyable : plus de 13 millions de personnes habitent dans cette ville! J'adore ça ! Hier on est allé sur la tour « Tokyo Skytree », Là, on a une magnifique vue sur toute la ville. C'était trop cool!

On mange souvent dans les restaurants : les sushi, « l'udon » (une soupe), et les glaces sont trop bons. Et il y a des automates avec des boissons ou même des « ramen » (des pâtes) partout. Et puis, aujourd'hui, à midi, on a mangé du « takoyaki » dans un restaurant. Ça n'avait pas l'air bon, mais quand j'ai essayé, j'ai changé d'avis : c'était délicieux ! Maintenant, c'est mon plat préféré !

On reste encore deux jours à Tokyo et après on va à Nara où il y a une statue en bronze gigantesque. Après Nara, on va à Okinawa. Là-bas, on a beaucoup de trucs à visiter. Moi, je veux aller au musée du karaté ! Je t'enverrai une photo.

On rentre en Suisse dans 10 jours. A bientôt! Timo

Huss rie stehn, weil es
rico darum geht

Links liest du einen Chat zwischen Timo und Martin.
Dazu stellen wir dir drei Fragen.

3. Frage:
Was will Timo in Okinawa machen?

- Im Restaurant essen.
- Eine grosse Statue sehen.
- Ein Museum besuchen.



THE LANGUAGE OF QUESTIONS AND RESPONSES: EVIDENCE ON ITS RELEVANCE

Qualitative evidence from verbal protocols

- **Switching** between languages was **not** perceived to be a problem (strongly confirmed by written questionnaire).
- **Questions** and multiple choice options in **French** were often **not** or **wrongly understood**.

I: ...Was magst du lieber?

S: Auf Deutsch. Weil beim Französischen ist man aufgeschmissen, wenn man die Frage nicht versteht.

Frage: Pourquoi le chemin de l'école de Vidal est-il spécial ?

I: Weisst du, wie die Frage lautet?
S: Warum ist das **Chemielabor** von Vidal so speziell.

I: ... What do you prefer?

S: In German. Because in French you're stuck when you don't understand the question.

Question: Why is Vidal's way to school special?

*I: Do you know what the question means?
S3: Why is Vidal's **chemistry lab** so special?*

Qualitative evidence from the interviews

- In French, **writing** was perceived as difficult.

S: Ich habe die Antwort, kann sie aber nicht auf Französisch schreiben.

S: I've got the answer but I can't write it in French.

- Many students **copied** their (mostly wrong) answer directly from the text.

I: War das Schreiben auf Französisch ein Problem?

S: Nein, das konnte ich ja abschreiben.

I: Was writing in French a problem?

S: No, I could just copy it.

Because of their writing difficulties, students had a tendency to copy from the French text passages that resembled more or less features of the (French) question.

Qualitative evidence from the interviews

- Students sometimes had answers in mind that were different from the ones they actually wrote down.

I: Kannst du mir deine Antwort kurz vorlesen?

S: Also Karusu verliert seinen Papa im Zoo.

I: Und geschrieben hast du „Karusu devenu papa“.

S: Ich habe geschrieben: “Pierre Dumont en danger” (...).

I: Was würdest du auf Deutsch schreiben (...)?

S: Pierre Dumont ist sehr gefährlich.

I: Can you read your answer?

S: Karusu loses his dad in the zoo.

I: And you wrote “Karusu devenu papa” [Karusu became a dad].

S: I wrote “Pierre Dumont en danger” [P. D. in danger] (...).

I: What would you write in German (...)?

S: Pierre Dumont is very dangerous.

Qualitative evidence from SAQ answers

Language of questions and answers	Absurd answers	Answers copied from text
French	33%	37%
German	10%	1%

Q: Où sont les trois personnes?

A: elle intéressant ce livre, imageset parler.

Q: *Where are the three people?*

A: *She interesting this book, [imageset] talk.*

Q: Qu'est-ce qu'Emilie préfère à l'école?

A1: Zurich.

A2: a monde.

A3: Il préféré à l'école jeudi.

Q: *What does Emilie like in school?*

A1: *Zurich.*

A2: *[a] world.*

A3: *He preferred at school Thursday.*

Q: Qu'est-ce que les deux veulent acheter?

A1: papa et garçon.

A2: et en plus tu n'aimes pas faire du vélo.

Q: *What do the two want to buy?*

A1: *dad and boy*

A2: *and you don't like to go by bike*

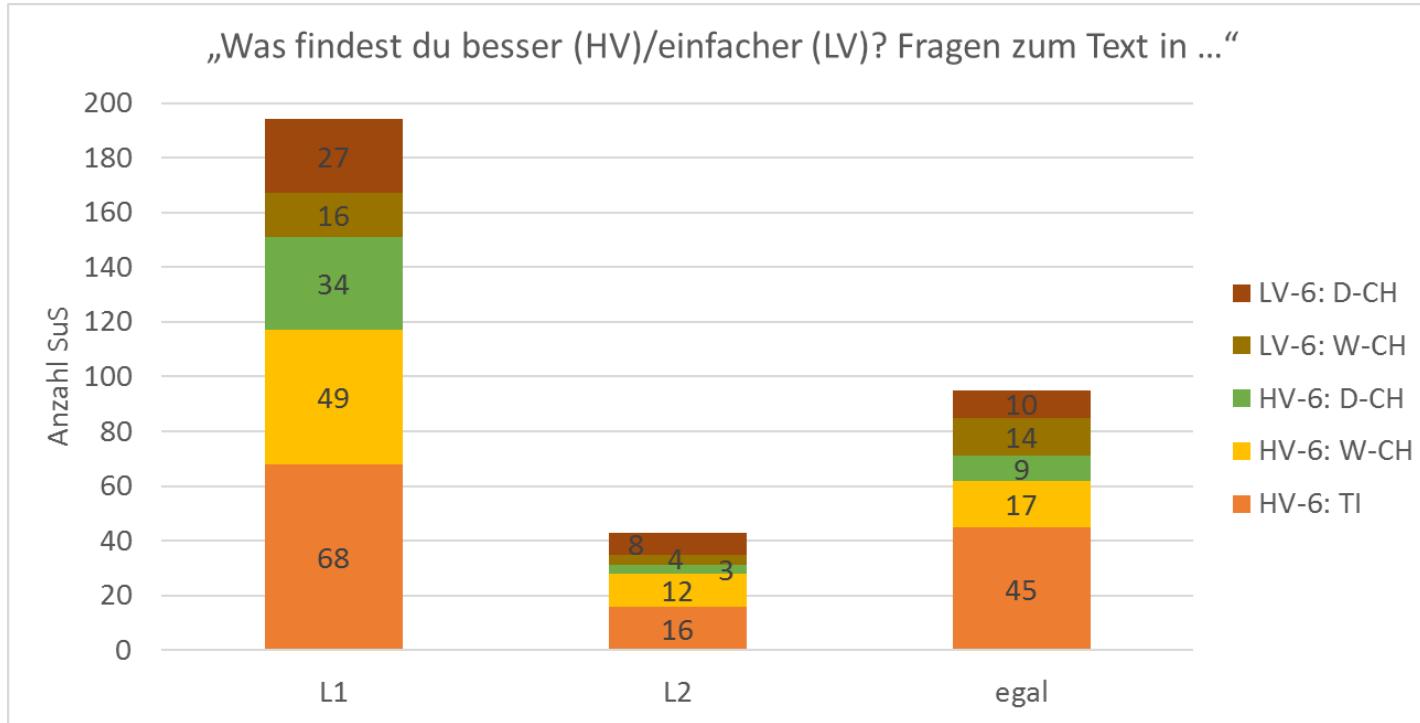
Evidence from questionnaires (I)

- Task Lab: 6th graders in the Passepartout region.
- Students had solved tasks in French with the questions and answers in either German or French.



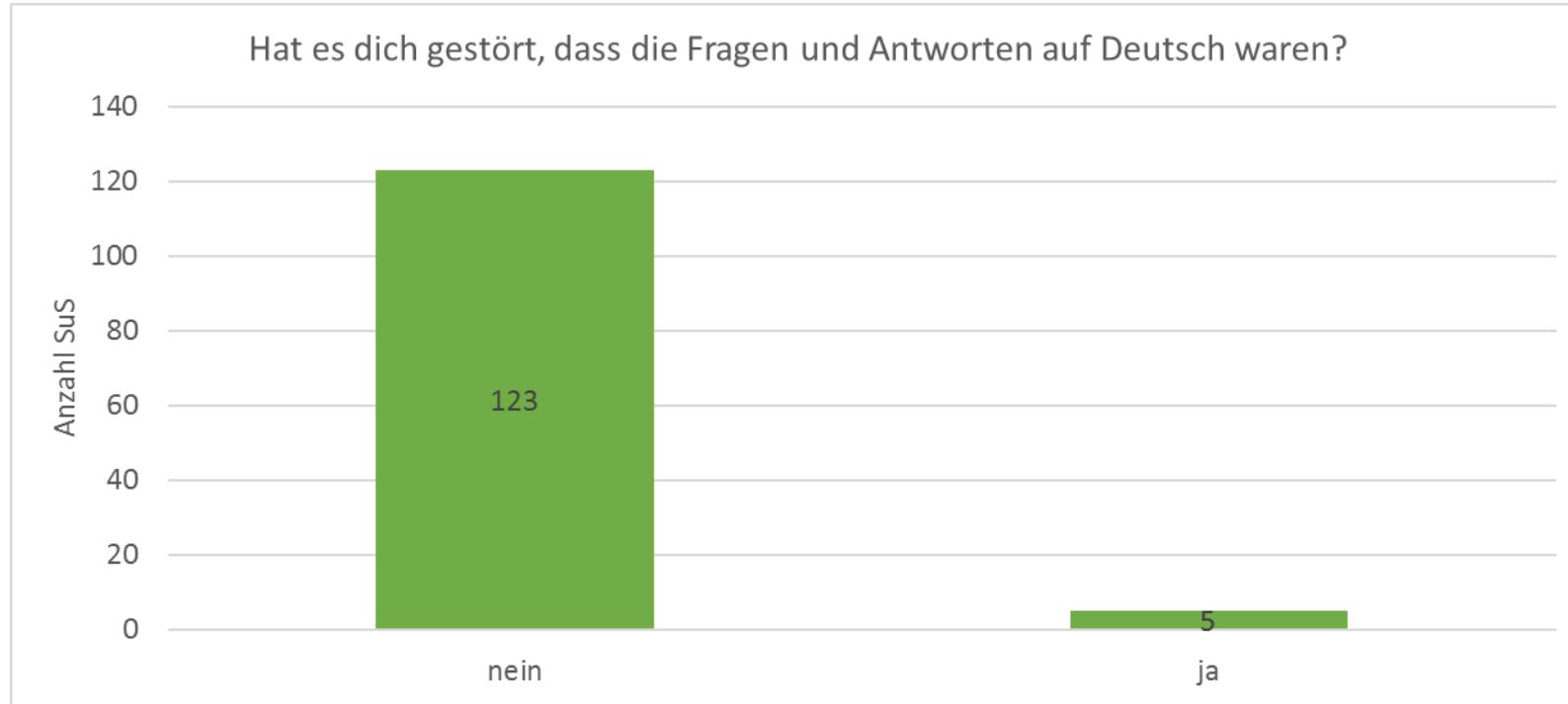
Evidence from questionnaires (II)

- Task development ÜGK: 6th graders in Ticino, Romandie & Passepartout.
- Students had solved tasks in their first foreign language (French or German) with the questions and answers in their L1 (Italian, French, German).



Evidence from questionnaires (III)

- IFB preliminary: 8th and 9ths-graders in German-speaking Switzerland.
- Students had solved tasks in either French or Englisch with the questions and answers in their L1.



Evidence from questionnaires: the reasons

L1 is better/easier/ok

Die L1 ist leichter zu verstehen.	<i>On comprend mieux!</i> <i>Perché è più semplice.</i> <i>Ich kann besser Deutsch als Englisch</i> <i>Parse que c'est ma langue naturel</i> <i>Percé è più comodo</i>
Fragen in der L1 versteht man sicher.	<i>Weil man vielleicht nicht alle fragen versteht. Und dann kann man nicht alles ausfüllen [...].</i>
Mithilfe der Fragen in der L1 kann man Inhalte des Textes erschliessen.	<i>ça m'aide à mieux comprendre le texte</i> <i>Perché se no non si capirebbe di cosa parla, [...] le domande ti aiutano.</i>
Texte in der L2 sind nicht verständlich.	<i>Weil ich nie Französisch spreche und fast nichts verstehe!</i> <i>Parce que je suis pas fort en allemand</i>

Evidence from questionnaires: the reasons

L2 is better/easier

Fragen und Antworten in der L2 können im Text wiedergefunden werden.	<p><i>Wenn der Texte und die Fragen auf Französisch sind, kann man das bestimmte Stichwort suchen und dann hat man die Antwort schon gefunden.</i></p> <p><i>Perché certe volte non si capisce bene e trovi già la risposta in Francese.</i></p>
HV: Geschriebene Wörter helfen beim Verstehen gesprochener Wörter.	<p><i>sono scritte nella stessa lingua in cui vengono pronunciate</i></p> <p><i>weil wenn er sagt j'mappel und da steht wie heisst er kann ich es besser wenn es Französisch.</i></p>
Mischung beider Sprachen stört, ist anstrengend oder ungewohnt.	<p><i>car sa perturbe</i></p> <p><i>Weil man zuerst überlegen musste was das auf Französisch heist.</i></p> <p><i>Parce que je suis abituer que les question sois en allemand</i></p>
Fragen in der L1 sind zu einfach.	<p><i>Weil es so einfach war</i></p> <p><i>Perché così ci si esercita un po' meglio il francese anche con le domande</i></p>

Evidence from questionnaires: the reasons

It's all the same to me

Man versteht beide Sprachen.	<p><i>Weil es schlussendlich nicht viel geähndert hatte, ich habe es trotzdem verstanden</i></p> <p><i>Capisco bene sia il francese sia l'italiano</i></p> <p><i>parceque j'arrive a comprendre assez les deux langues</i></p>
L2 ist zu schwierig, egal in welcher Sprache die Fragen sind.	<p><i>Ja weil ich nicht so gut bei Französisch bin darum ist es mir egal.</i></p> <p><i>je comprend pas l'allemand</i></p>
Beides hat Vorteile.	<p><i>En français on sais comme c'est ma langue maternelle. En allemand on repère les mots.</i></p>
Es wäre eine Herausforderung.	<p><i>[...] wenn die Fragen auf Französisch gewesen wären, wäre dass eine Herausforderung gewesen → das wäre cool.</i></p> <p><i>C'est plus facile en français mais se serait encore plus drôle et pimenté de mettre des questions en allemand</i></p>

Some sample analyses of the *Task Lab* items

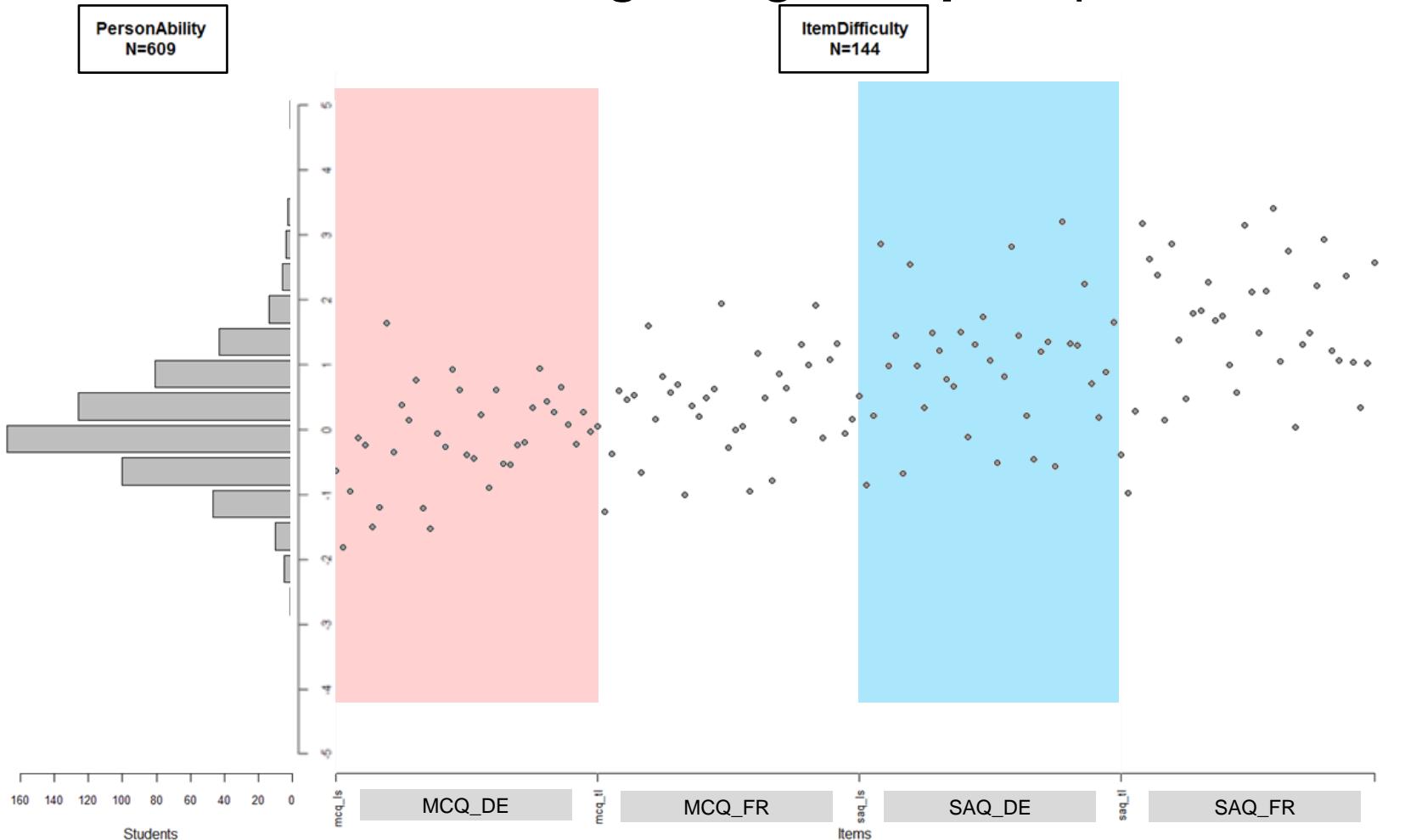
...focusing on the two format aspects *language* and *item type* (*test method*)

		Language of Qu.&Re	
		DE	FR
Item Type	MCQ	MCQ_DE	MCQ_FR
	SAQ	SAQ_DE	SAQ_FR

N items: 4 x 36 such items (on 12 different reading passages)

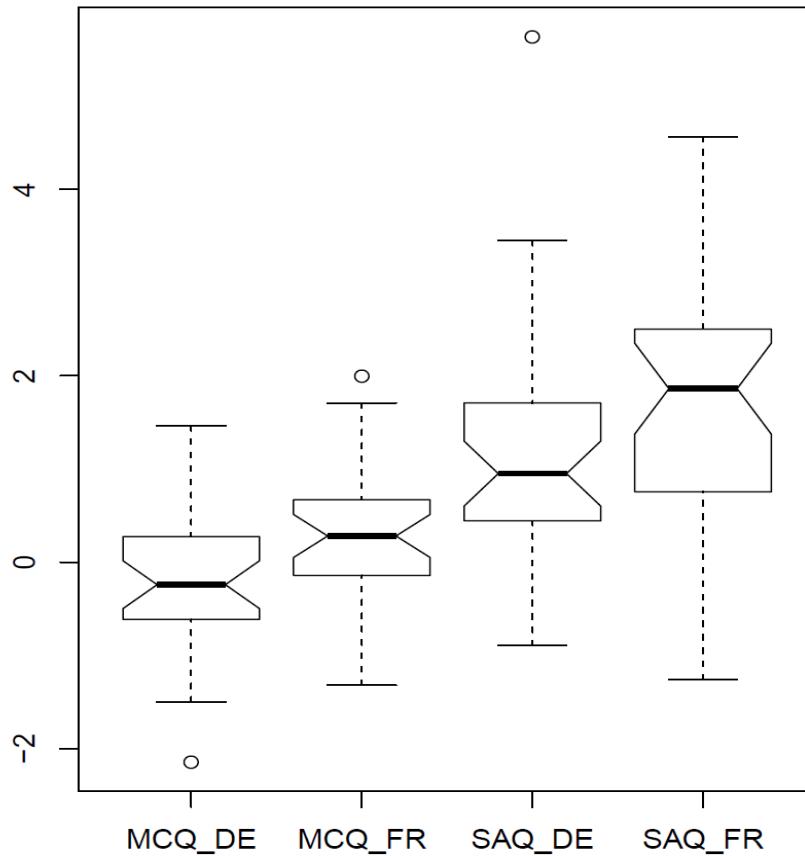
N pers.: 35 classes, 609 students (\varnothing 120 students per reading task)

Result of Rasch scaling: Wright Map for persons & items



Item difficulty per item group

Item Difficulties on Item Type X Language



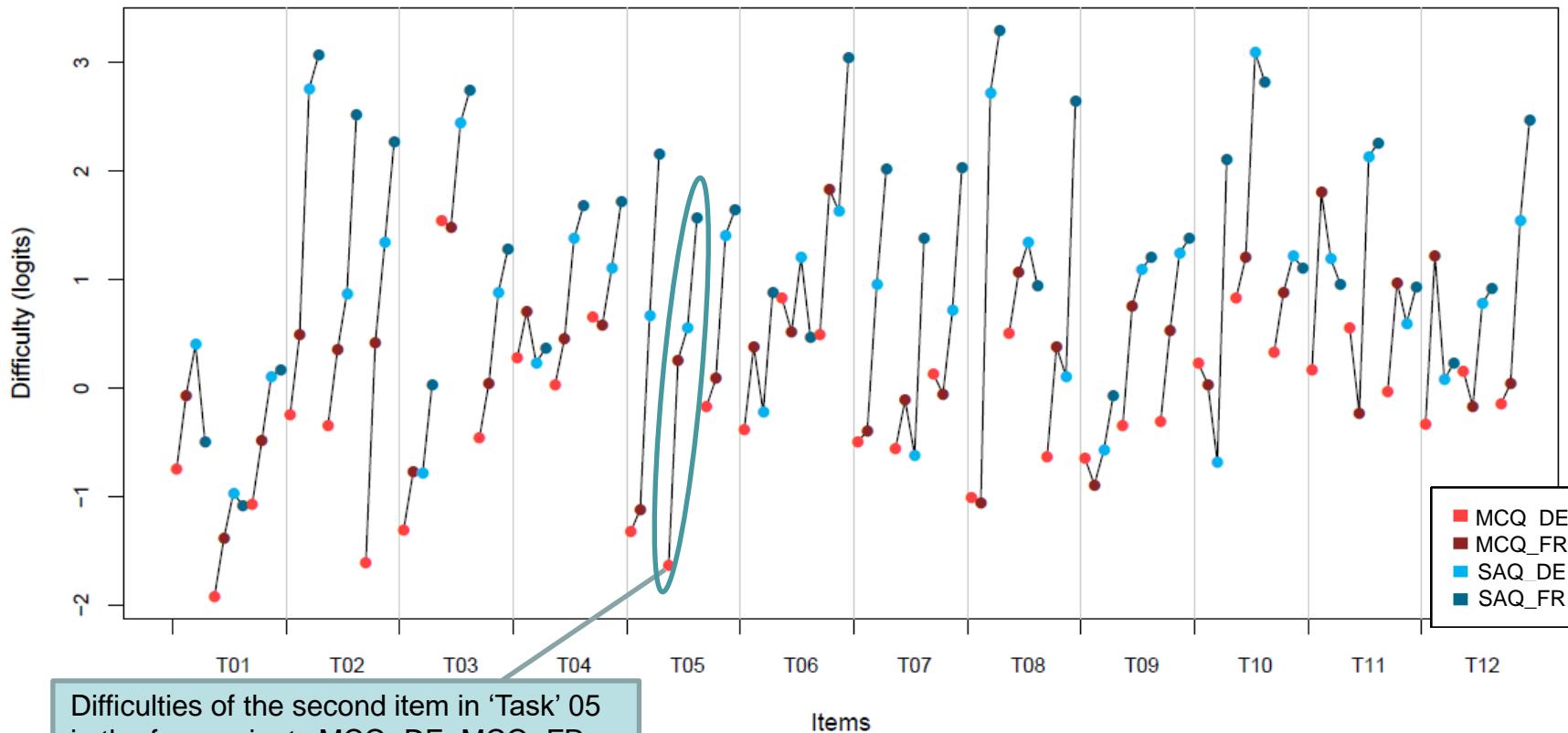
Should this bother us?

- Possibly: "same items" are not equal in difficulty
 - There are factors around we don't know, i.e. which are even undesirable
- => A detailed look may help understand and improve item construction

Difficulty of individual items per item group (I)

Overview

Item difficulty 1-PL
by format and language



Difficulty of individual items per item group (I)

Overview

- “Standard pattern”: MCQ is easier than SAQ and German is easier than French (13 items)
 $MCQ_DE < MCQ_FR < SAQ_DE < SAQ_FR$
- Three major deviations:
 - MCQ_FR is easier than MCQ_DE (7 items)
 $MCQ_DE > MCQ_FR$
 - SAQ_FR is easier than SAQ_DE (5 items)
 $SAQ_DE > SAQ_FR$
 - SAQ_DE is easier than MCQ_FR (8 items)
 $MCQ_FR > SAQ_DE$
- Combinations of the deviations (1 item each)

} French is easier than German

Difficulty of individual items per item group (II)

“Standard pattern”: MCQ easier than SAQ, German easier than French (13 items)

Une conversation

Personnes: A, B et C

A : « Mais ce n'est pas intéressant ce livre ! »

B : « Pourquoi ? J'aime bien les images. »

A : « Mais il n'y a pas beaucoup à lire ! 3 m une page. Je lis ça en 30 minutes. »

B : « Ah, moi, j'adore. J'en prends deux ou prends quoi ? Il a combien de pages ton livr

A : « Mais non, 350. Regarde, c'est « Les aventures de Sherlock Holmes ». C'est super ! »

C : « Vous faites quoi là ? Je vous demande de rester silencieux. On vient ici pour lire ou pour faire ses devoirs, pas pour parler ! »

A : « Ah, excusez-nous, Madame. On pensait qu'on venait ici uniquement pour chercher des livres. »

C : « Ah, mais non ! Bien sûr que vous pouvez emprunter des livres. Mais vous pouvez aussi venir ici pour lire ou pour faire vos devoirs après l'école. On a aussi des ordinateurs si vous voulez chercher quelque chose sur Internet. Mais pour ça, il faut du silence. »

B : « Okay, la prochaine fois, on ne va pas pa

C : « Parfait. Alors, vous voulez prendre des li

Item in FR depends on understanding “où” and “qui”.

1. Où sont les trois personnes?

- Dans une salle d'ordinateurs.
- Dans une salle de classe.
- Dans une bibliothèque.

2. Qui sont A et B?

- Ce sont des élèves.
- Ce sont des professeurs d'école.
- Ce sont des bibliothécaires.

3. Quel est le problème avec C dans cette scène?

- A et B doivent faire leurs devoirs.
- A et B ne doivent pas parler.
- A et B ont perdu un livre.

No answer option can be found verbatim in the text.
SAQ cannot be answered by citing passages.
→ Relevant text passages must be understood!

T05

Difficulty of individual items per item group (III)

French is easier than German (8 items MCQ, 6 items SAQ, 1 item both)

MCQ_DE > MCQ_FR | SAQ_DE > SAQ_FR

The answer can be found verbatim in the text.
Distractors can be easily identified as wrong.

Tom écrit un texte pour l'école

Bonjour. Je m'appelle Tom et j'ai 13 ans. J'ai un grand frère de 17 ans et une petite sœur de 8 ans. Nous habitons à la campagne dans une grande ferme.

Ma mère est vétérinaire et mon père est paysan comme mes grands-parents : ils travaillent dans la ferme et s'occupent des animaux. C'est pourquoi nous habitons dans la même maison que mes grands-parents. Il y a aussi mon oncle. Maintenant, il est à la retraite, mais avant, il était professeur de physique.

Pendant la journée, le matin et l'après-midi, chacun travaille ou est à l'école et on ne se voit pas beaucoup. A midi, nous, les enfants, restons à l'école. Mais chaque soir, nous mangeons tous ensemble à la maison. Chacun parle de sa journée. J'aime beaucoup ça.

J'adore ma famille, mais parfois je m'énerve un peu : tout le monde parle beaucoup et fort et c'est le grand chaos !

1. Quel est le thème du texte de Tom?

- La vie à la campagne.
- La vie des animaux de la ferme.
- La vie dans sa famille.

2. Quel travail font les grands-parents de Tom?

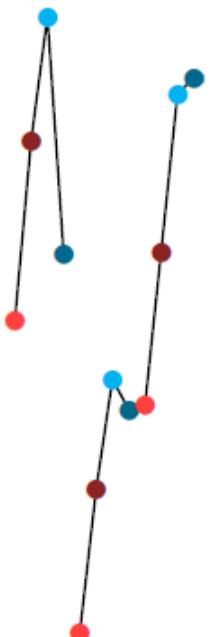
- Ils sont vétérinaires.
- Ils sont paysans.
- Ils sont de professeurs de physique.

3. Quand est-ce que toute la famille se voit?

- Le matin.
- A midi.
- Le soir.

Difficulty of individual items per item group (III)

French is easier than German (8 items MCQ, 6 items SAQ, 1 item both)



T01

Emilie se présente

Je m'appelle Emilie. J'ai un petit frère, Maxime, et une grande sœur, Caroline. J'habite à Renens.

C'est une petite ville à côté de Lausanne. Mon frère a un petit chat, il s'appelle Philippe. Il est noir et blanc et il a le même anniversaire que le copain de Caroline !

A l'école, j'aime la géographie. Je suis intéressée par tous les pays du monde. Je n'aime pas la biologie, parce que c'est vraiment compliqué. Ma copine Sarah préfère l'allemand, parce que son père vient de Zurich.

Chaque jeudi après l'école, je vais à Lausanne pour un cours de théâtre. Je trouve ça trop cool ! J'adore aussi jouer aux jeux-vidéos et passer du temps avec mes copines, faire du sport ou jouer dehors.

The answers can be copied from the text. The last answer can easily be translated into German.

1. Qui est Philippe ?



2. Qu'est-ce qu'Emilie préfère à l'école ?

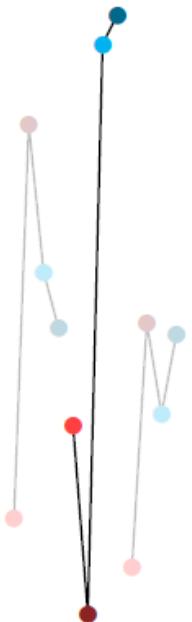


3. Qu'est-ce qu'Emilie fait à Lausanne ?



Difficulty of individual items per item group (III)

French is easier than German (8 items MCQ, 6 items SAQ, 1 item both)



The answer can be found verbatim in the text.
Distractors can be easily identified as wrong.

The correct answer – if understood – is counter-intuitive.

Ça va ? Moi je vais très bien . Mon test d'anglais d'hier a été annulé .
Je te raconte ce qui s'est passé hier. C'était le 1er avril. En Suisse, le 1er avril, on joue des tours aux autres.
On appelle ça « faire des poissons d'avril ».
Hier, on m'a joué beaucoup de tours : le matin, quelqu'un a changé l'heure de mon réveil : il a sonné à 4 heures du matin... Après, j'ai mis du lait dans mes céréales. Et il était tout vert. Beurk ! Puis, avant d'aller à l'école, il y avait du papier toilette dans mes chaussures...
Et, finalement, la meilleure blague du jour : quand je suis arrivé à l'école, on avait tous peur du test d'anglais. Mais heureusement, le professeur nous a dit que c'était un poisson d'avril ! Ouf !! Je n'avais pas beaucoup étudié !
Et toi ? Quelqu'un t'a aussi fait des blagues comme ça ? En tout cas, dans un an, ça sera moi qui jouerai des tours aux autres : attention tout le monde ! J'ai toute une année pour inventer les meilleures blagues. J'ai déjà la première idée : je vais échanger le sucre et le sel ! Qu'est-ce que tu en penses ?

A bientôt,
Samuel

1. Quel est le thème du mail de Samuel ?

- L'école en Suisse.
- Le test d'anglais du jour d'avant.
- Les blagues qu'on lui a faites.

2. Quelle était la blague préférée de Samuel ?

- La blague préférée de Samuel.
- Le test d'anglais.
- Le sel dans le thé.
- Le papier toilette dans les chaussures.

3. Qu'est-ce que Samuel fera le 1er avril prochain ?

- Il ne sait pas encore. Il demande à Benjamin quoi faire.
- Il va mettre le sel dans la boîte à sucre.
- Il va faire attention aux blagues d'autres.

Difficulty of individual items per item group (III)

French is easier than German (8 items MCQ, 6 items SAQ, 1 item both)

The correct answer – if understood – is counter-intuitive.

Les personnes: le Reporter (50 ans), Alice (15 ans) et Simone (15 ans)

R : « Bonjour. Ma première question : quand vous voulez parler avec vos copains, qu'est-ce que vous faites ? »

A : « Moi j'aime beaucoup parler avec mes copines après l'école. Quand je suis à la maison, je leur écris des messages. »

R : « Et comment tu leur écris ? »

A : « Avec mon Natel. A l'école, on s'écrit aussi des petits papiers, parce que les téléphones sont interdits. »

R : « Et toi, Simone ? »

S : « Moi, je n'aime pas trop écrire. »

R : « Alors tu téléphones avec tes copines ? »

S : « Oui, mais ce sont des appels très courts, une ou deux minutes. On prend le Natel juste pour s'organiser ou pour se dire à quelle heure on se voit pour faire du sport. »

R : « Et toi Alice, tu téléphones souvent à tes amies ? »

A : « Oh oui ! J'appelle mes copines tous les jours, même si on se voit à l'école. Nous, on peut se parler tout le temps ! »

R : « La communication entre les jeunes a bien changé. Ça va plus vite maintenant, et c'est bien comme ça. C'est dommage que je n'avais pas de Natel quand j'étais jeune. Mais maintenant, je vais m'acheter un smartphone tout de suite.

Merci beaucoup à vous deux ! Au revoir. »

S+A : « Au revoir. »

T07

1. Warum macht der Reporter dieses Interview?

Der Reporter interessiert sich...

- für die Schule von heute.
- für die Kommunikation von Jugendlichen heute.
- für die Freunde der Jugendlichen von heute.

2. Wer benützt sein Natel am häufigsten?

- Simone.
- Der Reporter.
- Alice.

3. Was denkt der Reporter nach dem Interview über Natels?

Der Reporter findet, dass...

- die Leute ihre Natels nicht zu viel benutzen sollten.
- die Natels praktisch sind, um mit den Freunden zu sprechen.
- die Schulen die Natels verbieten sollten.

Difficulty of individual items per item group (IV)

SAQ_DE is easier than MCQ_FR (10 items)

MCQ_FR > SAQ_DE

Tom écrit un texte pour l'école

Bonjour. Je m'appelle Tom et j'ai 13 ans. J'ai un grand frère de 17 ans et une petite sœur de 8 ans. Nous habitons à la campagne dans une grande ferme.

Ma mère est vétérinaire et mon père est paysan comme mes grands-parents : ils travaillent dans la ferme et s'occupent des animaux. C'est pourquoi nous habitons dans la même maison que mes grands-parents. Il y a aussi mon oncle qui vit à la retraite, mais avant, il était professeur de français.

Pendant la journée, le matin et l'après-midi, je vais à l'école et on ne se voit pas beaucoup. Nous nous retrouvons tous ensemble à la maison. Chacun parle de ce qu'il a fait pendant ça.

J'adore ma famille, mais parfois je m'énerve. Parce que le monde parle beaucoup et fort et c'est le cas.

1. Quel est le thème du texte de Tom?

- La vie à la campagne.
- La vie des animaux de la ferme.
- La vie dans sa famille.

2. Quel travail font les grands-parents

In French MCQ they tend to (wrongly) match words from the options.

In German SAQ students are able to identify the general subject of a text using familiar vocabulary.

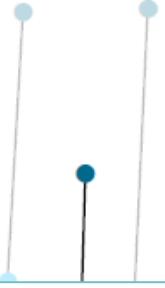
„Also oben erzählt er ein bisschen mehr über sein Leben, und unten über die Schule. Und dann wusste ich, um was es eigentlich geht.“

„Er stellt sich vor, der Tom.“

T06

Difficulty of individual items per item group (IV)

SAQ_DE is easier than MCQ_FR (10 items)



Une interview

Les personnes: le Reporter (56 ans), Alice (13 ans) et Simone (13 ans)

R : « Bonjour. Ma première question : quand vous voulez parler avec vos copains, qu'est-ce que vous faites ? »

A : « Moi j'aime beaucoup parler avec mes copines après l'école. Quand je suis à la maison, je leur écris des messages.»

R : « Et comment tu leur écris ? »

„Ich bin auf die Idee gekommen, dass Alice es am meisten braucht. Den Reporter und Simone habe ich ausgeschlossen.“

I: Was hast du in der Frage verstanden?

26: „natel“. Weil „utilise“ weiss ich gerade nicht.

I: [...] Was hast du dir gedacht, wie die Frage lauten könnte?

26: Warum das Natel so „souvent“.

Although students can find the answer in the German version, they don't understand the French question.

1. Pourquoi le reporter fait cette interview ?

Le reporter s'intéresse...

- à l'école d'aujourd'hui.
- à la communication entre les jeunes d'aujourd'hui.
- aux amis des jeunes d'aujourd'hui.

2. Qui utilise son Natel le plus souvent?

- Simone.
- Le reporter.
- Alice.

3. Que pense le reporter des Natels après l'interview?

Il pense que...

- les gens ne devraient pas trop utiliser les Natels.
- les Natels sont pratiques pour parler avec les copains.
- les écoles doivent interdire les Natels.

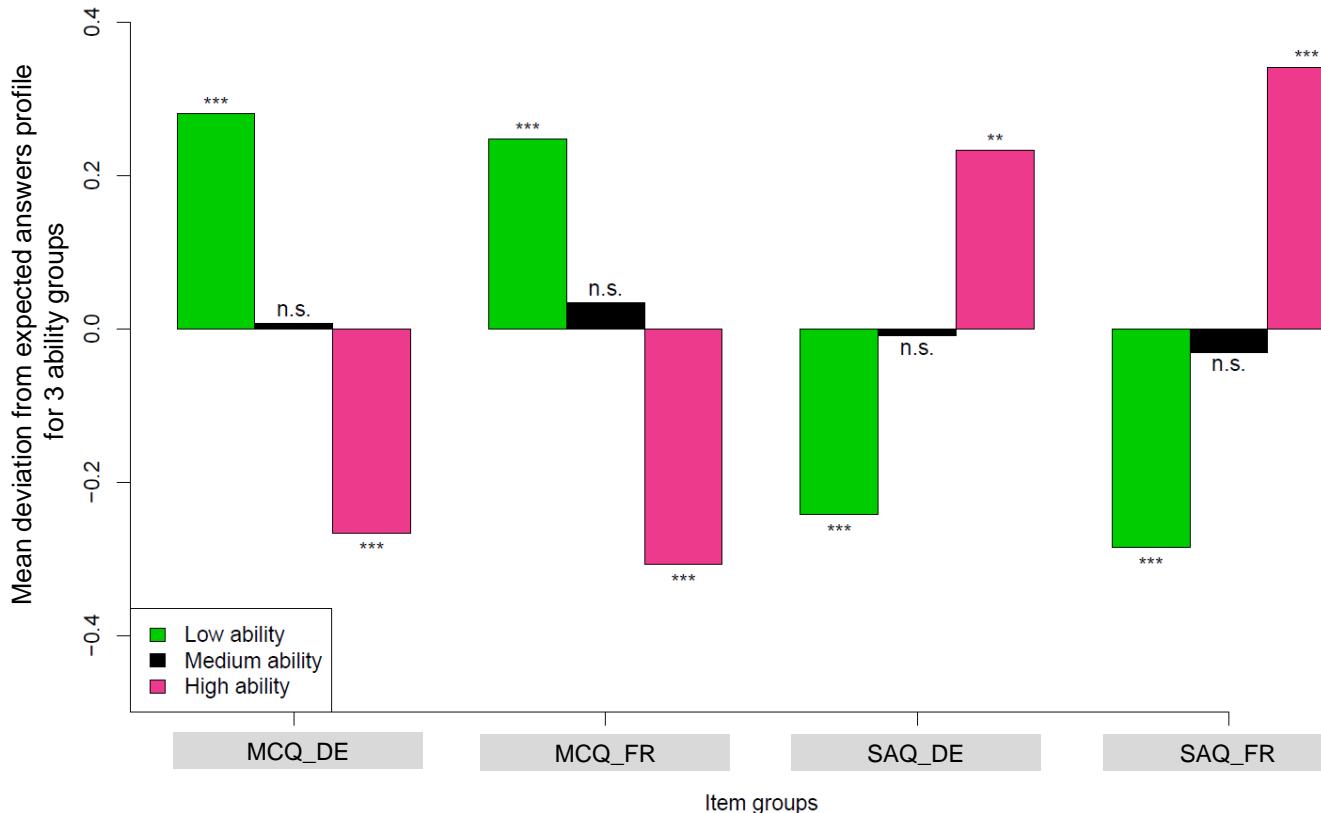
In sum:

In our context (level etc.), L2 for questions and responses in a reading test poses validity problems

- Proper *text* reading doesn't take place when the question is unclear.
- In French MCQ matching corresponding words in question/options and text seems to be the default strategy when in trouble.
- In French SAQ answering strategies are motivated by the need to write in L2. Students copy available text elements.

Item difficulty per item group: *Profile Analysis*

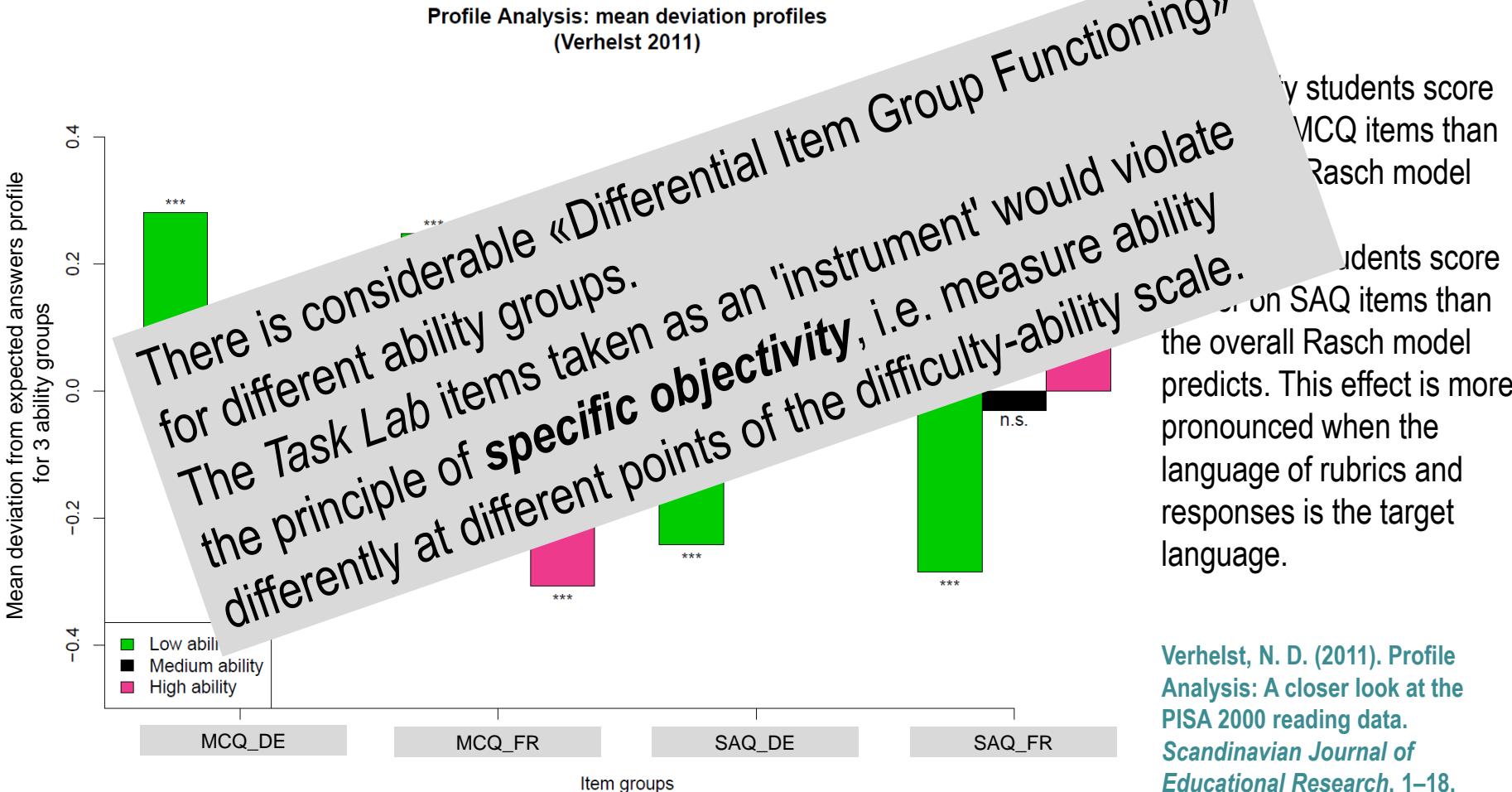
Profile Analysis: mean deviation profiles
(Verhelst 2011)



- Low-ability students score higher on MCQ items than the overall Rasch model predicts.
- High-ability students score higher on SAQ items than the overall Rasch model predicts. This effect is more pronounced when the language of rubrics and responses is the target language.

Verhelst, N. D. (2011). Profile Analysis: A closer look at the PISA 2000 reading data. *Scandinavian Journal of Educational Research*, 1-18.

Item difficulty per item group: *Profile Analysis*



Item fit analysis – a heuristic tool for quality assess.

Two well-known (but not uncontroversial) statistics:

Outfit and Infit Mean Square.

Outfit mean square summarizes the difference between the observed person scores on an item and the person scores predicted by the model on that same item.

The **infit mean square** weights extreme differences (e.g. very weak student succeeds on a very hard item) less than differences in the central region. It is often preferred. **Infit > 1.2** may indicate a problem (> 15% of unexpected variability relative to the model). For more precise information check **ICC!**

	parameter	Infit	Infit_t	Infit_p	Infit_phot
1	lv.saq_T01_1_ls	0.9091704	-1.6102363454	0.1073462697	1.0000000
2	lv.saq_T01_2_ls	0.9982843	-0.0011942772	0.9990471049	1.0000000
3	lv.saq_T01_3_ls	0.9009748	-1.9418915319	0.0521502328	1.0000000
4	lv.saq_T02_1_ls	0.8199686	-0.5014013947	0.6160886584	1.0000000
5	lv.saq_T02_2_ls	1.0371765	0.3794236195	0.7043733137	1.0000000
6	lv.saq_T02_3_ls	0.8867375	-0.8128006135	0.4163323837	1.0000000

Visualizing item fit: the Item Characterstic Curve

Low discrimination, underfit & misfit

High discrimination: overfit

Un mail de Samuel

Links siehst du ein Mail, das Samuel an seinen Freund Benjamin schreibt.

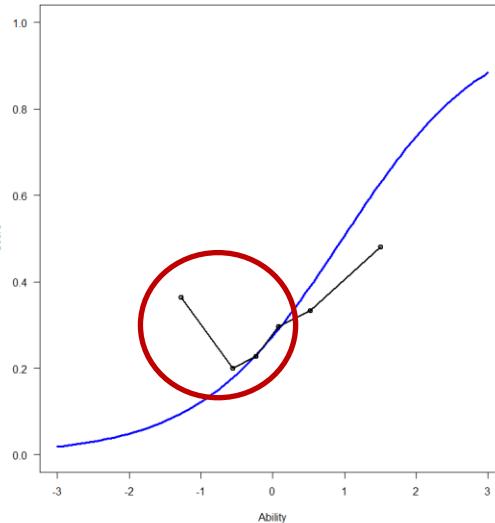
Kreuze die richtige Antwort an.

De : Samuel
A : Benjamin
Date : 2 avril, 8h15
Objet : Hello from Lausanne !

Salut Benjamin !

Ça va ? Moi je vais très bien : mon test d'anglais d'hier a été annulé !
Je te raconte ce qui s'est passé hier. C'était le 1er avril. En Suisse, le 1er avril, on jor

Expected Scores Curve - Item Iv.mpc_T11_3_tl



1. Quel est le thème du mail de Samuel ?

- L'école en Suisse.
- Le test d'anglais du jour d'avant.
- Les blagues qu'on lui a faites.

2. Quelle était la blague préférée de Samuel ?

- Le test d'anglais.
- Le sel dans le thé.
- Le papier toilette dans les chaussures.

3. Qu'est-ce que Samuel fera le 1er avril prochain ?

- Il ne sait pas encore. Il demande à Benjamin quoi faire.
- Il va mettre le sel dans la boîte à sucre.
- Il va faire attention aux blagues d'autres.

Un chat entre Timo et Martin

Martin : Salut Timo ! Ça va les vacances ? Toujours dans la grande capitale japonaise ? A bientôt, Martin.

Timo: Salut ! Oui, on est toujours là, c'est super !!! C'est vraiment incroyable : plus de 13 millions de personnes habitent dans cette ville ! J'adore ça ! Hier on est allé sur la tour « Tokyo Skytree ». Là, on a une vue magnifique sur toute la ville. C'était trop cool !

On mange souvent dans les restaurants : les sushi, les « ramen » (des pâtes), et les glaces sont trop bons. Et il y a des automates avec des boissons ou même « l'udon » (une soupe) partout. Et puis, aujourd'hui, à midi, on a mangé du « takoyaki » dans un restaurant. Ça n'avait pas l'air bon, mais c'était délicieux ! Maintenant,

On reste encore deux jours à une statue en bronze gigante. Là-bas, on a beaucoup de tru du karaté ! Je t'enverrai une photo. On rentre en Suisse dans 10

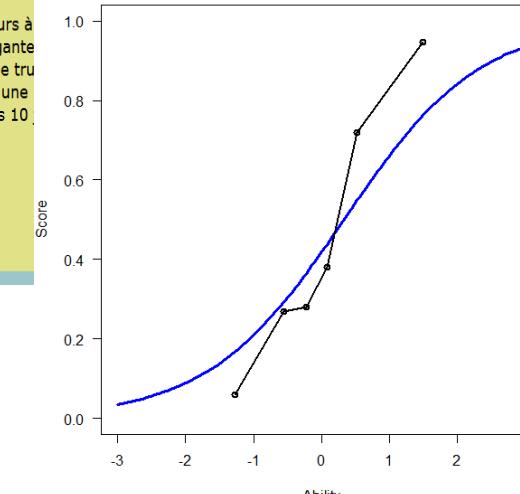
Links siehst du einen Chat zwischen Timo und Martin.
Dazu stellen wir dir drei Fragen.

1. Frage:

Was ist das Thema von Timos Nachricht?

Schreibe deine Antwort auf Deutsch!

Expected Scores Curve - Item Iv.saq_T04_1_tl



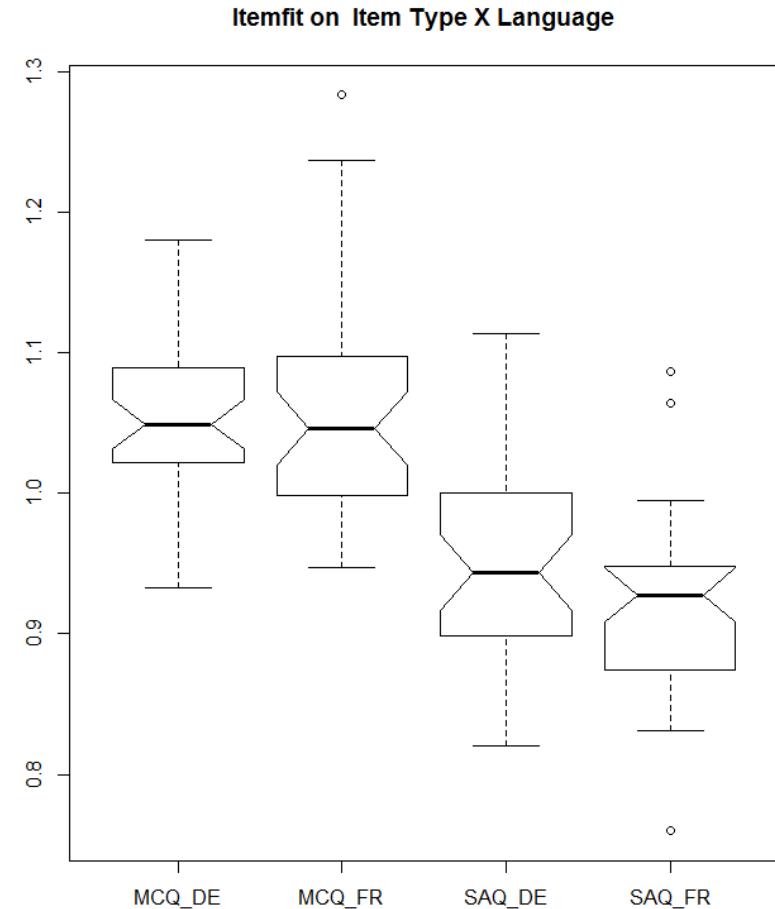
Misfit analysis if different item groups are present as in our case

General tendencies:

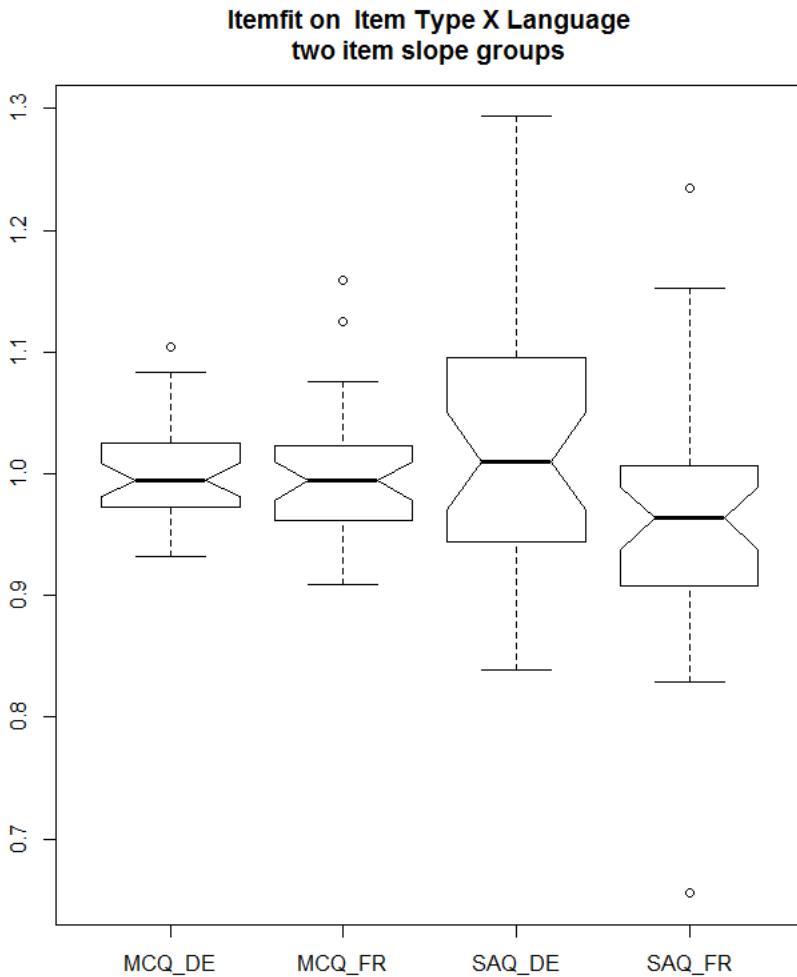
MCQ items generally have infit values > 1 (*underfit*) because they discriminate less.

SAQ items generally have infit values < 1 (*overfit*) because they discriminate more.

=> **Be careful:** Infit values are misleading!
Carry out **separate** fit analyses for
2 item groups or **define slope groups**



Misfit analysis based on two different slope groups



When two slope groups are defined, the infit statistic can be used to detect problems causing misfit.

Check items with *high infit*. It is usually a sign that *construct-irrelevant variance* has an influence.

On the low infit side, only very low fit values are worth inquiring.

Detecting misfit in items

	parameter	Infit	Infit_t	Infit_p	Infit_pholm
2	lv.saq_T01_2_ls	1.1034843	1.1016794	0.2706010881	1.00000000
5	lv.saq_T02_2_ls	1.1767505	1.4471564	0.1478531316	1.00000000
7	lv.saq_T03_1_ls	1.1209062	1.2526753	0.2103238886	1.00000000
16	lv.saq_T06_1_ls	1.1538618	1.9649232	0.0494231275	1.00000000
17	lv.saq_T06_2_ls	1.1642487	1.2992073	0.1928728133	1.00000000
24	lv.saq_T08_3_ls	1.2933722	3.4302113	0.0006031114	0.08684804
27	lv.saq_T09_3_ls	1.1859465	1.2954973	0.1951487337	1.00000000
31	lv.saq_T11_1_ls	1.1957725	1.5111295	0.1307554677	1.00000000
58	lv.saq_T08_1_t1	1.1519310	0.4698663	0.6384505627	1.00000000
67	lv.saq_T11_1_t1	1.2343125	2.0947628	0.0361920814	1.00000000
106	lv.mpc_T06_2_t1	1.1588436	2.1831165	0.0290272283	1.00000000
126	lv.mpc_T09_3_t1	1.1245595	2.3171081	0.0204978457	1.00000000
139	lv.mpc_T12_1_ls	1.1045118	1.9067241	0.0565563227	1.00000000
54	lv.saq_T06_3_t1	0.6553845	-1.3290684	0.1838254179	1.00000000

>1 reasonable answer?

A technical solution for DIGF: a 2PL IRT model

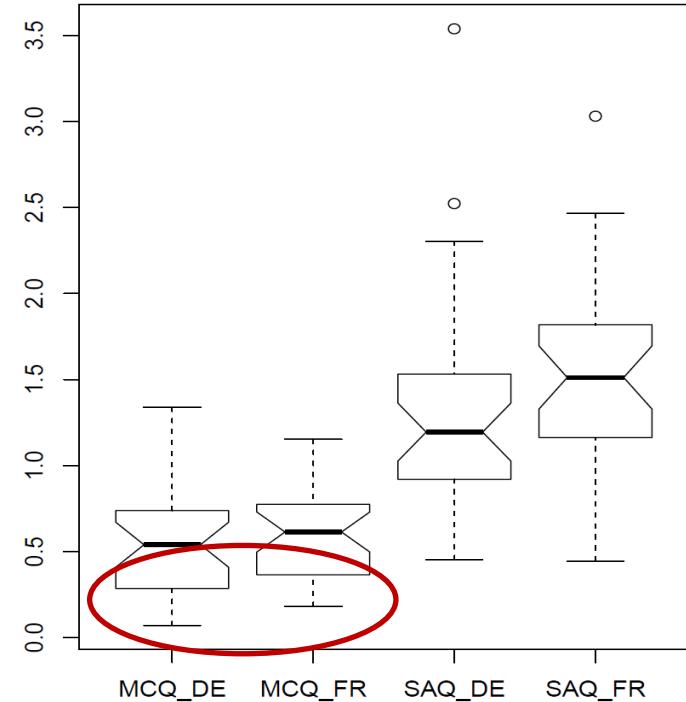
The items of the four groups don't contribute equally to the measurement of the construct

A 2PL IRT model takes this into consideration by **estimating a slope parameter (discrimination) for the individual items.**

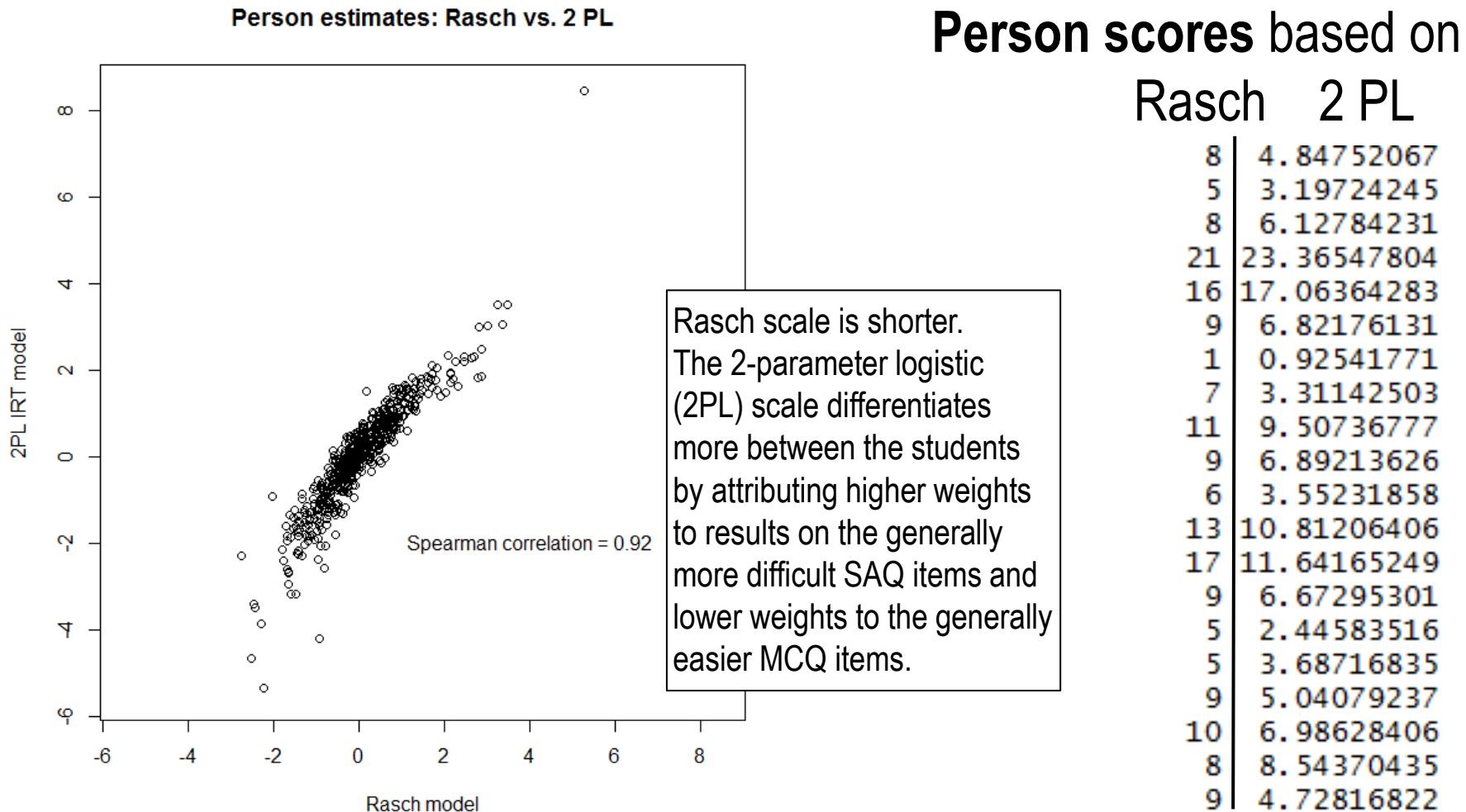
=> Discrimination of MCQ items is generally much lower.
Some items hardly separate generally strong from generally weak students

=> 2PL model weights the scores on the items according to discrimination.

Item Discriminations on Item Type X Language



Person (WLE) measures: Rasch vs. 2 PL models



DOES SUCCESS ON MCQ AND SAQ HAVE THE SAME PREDICTORS? – LATENT REGRESSION ON 2 DIMENSIONS

A 2-dimensional (per item type) model fits better than the 1-dimensional model

Mod. 2PL.1Dim vs. Mod. 2PL.2Dim (dev. Diff.) : $\chi^2(10.85, 1)$, p < 0.001

Results of latent ('error-free') regression on 2 dimensions:

Predictors	Predictor type	Dim 1 SAQ	Dim 2 MCQ
Gender: male	dummy	0.095	-0.115
Romance lang. background	dummy	0.512	0.112
Motivation: enjoyment	z-std.	0.181	0.109
Motivation: ought	z-std.	0.038	-0.013
Backward digit span	z-std.	0.150	0.145
Sight-word recognition	z-std.	0.159	0.145
Yes-No Test (recognise word)	z-std.	0.142	0.292
Segmentation task	z-std.	0.406	0.293
C-Test	z-std.	0.221	0.156

Predictors	Predictor type	Dim 1 SAQ	Dim 2 MCQ
Gender: male	dummy	0.095	-0.115
Romance lang. background	dummy	0.512	0.112
Motivation: enjoyment	z-std.	0.181	0.109
Motivation: ought	z-std.	0.038	-0.013
Backward digit span	z-std.	0.150	0.145
Sight-word recognition	z-std.	0.159	0.145
Yes-No Test (recognise word)	z-std.	0.142	0.292
Segmentation task	z-std.	0.406	0.293
C-Test	z-std.	0.221	0.156

Some Observations

- Known **correlates** of better language knowledge predict success on **SAQ** items *particularly well*: Integrative measures, Motivation (enjoyment ≈ intrinsic), and a romance family language background (13.6% of sample).
- The strictly **receptive** Yes-No word recognition test is a better predictor for success on **MCQ**. The possibility of success through **guessing** may be a commonality (despite a correction for guessing made on YNT).

Discussion

- It seems desirable to be able to pinpoint more **specific component knowledge & skills – which ones?**
- What **item characteristics** should be taken into account for a rigorous **person-item explanatory model?**

TAKING THINGS FURTHER: THE *IFB* PROJECT

Innovative Forms of Assessment (IFB): Goals (I)

- Development of “innovative” assessment tasks (potentially) for large-scale assessment in **9th grade**
 - **Scenario-based** assessment (SBA): Overarching topic, integrated reading and listening
 - “Modern” text types (ICT), suitable for the target group
 - Computer-based administration
 - For French and English as foreign languages

Innovative Forms of Assessment (IFB): Goals (II)

- Validation of the SBA tasks using qualitative and quantitative methods
 - Construct validation, e.g. cognitive process validity
 - Correlation with known **components** of reading and listening ability (language-specific, strategies and IQ)
 - Correlation with established forms of reading and listening assessment (**concurrent validity**)
 - Correlation with background variables (e.g. SES)

IFB: glimpses of the scenario "A Trip to Limerick"



Du hast Kollegen in Limerick: Eddie, Harry und Anne.

In einem Gruppenchat geben sie dir Tipps, was man in Limerick unternehmen kann.

Auf den nächsten Seiten sammelst du die Infos deiner Kollegen in einer Liste.

Weiter

Introductory page: overview of setting and task

IFB: "Trip to Limerick" (part 1)

The screenshot shows a mobile messaging application interface. At the top, it says "The Limerick Group" and lists participants: You, Harry, Eddie, Anne. The conversation starts with Eddie: "Hey guys! Our teacher told us about Limerick Castle today. Should we go there?". Anne responds: "The castle is a classic and a number one sight in town. But honestly? I wouldn't go there." Harry follows up: "No, you would be bored for hours! Limerick has a new sports stadium though, you could go there!". Eddie asks: "You shouldn't take a tour of the stadium itself, but there is a rugby museum inside. You have to visit that. It's really cool." Anne replies: "Nice. Does it take long to visit?". Eddie answers: "No, an hour or two. After that, you can get some great lunch at the Red Hen. Coolest restaurant in town!". Eddie concludes: "Yeah! You should try their burgers!". Below the messages are standard mobile navigation icons: back, home, and recent apps.

Lies den Anfang des Chats.

Ziehe die Informationen in die passende Kategorie.
Die Reihenfolge ist egal.

Das sollten wir machen:	Das sollten wir nicht machen:
?	?
?	?

Das Schloss anschauen

Das Stadion besichtigen

Die Rugby-Ausstellung besuchen

Im Restaurant 'Red Hen' essen

Weiter

Reading: identify recommendations

IFB: "Trip to Limerick" (part 2)

Höre den ersten Teil von Lous Podcast. Beantworte dann die Frage.

Lou war für drei Tage nicht im Internet. Warum?

- Sie war drei Tage in einem abgelegenen Dorf am Meer.
- In ihrem Hotel in Limerick funktionierte das Internet nicht.
- Sie musste warten, bis ihr Smartphone repariert wurde.

Weiter

Listening: travel blog

IFB: "Trip to Limerick" (part 3)

The screenshot shows a web browser window with the URL <https://www.hostelo.com/Limerick/The-Limerick-Hotel>. The page title is "The Limerick Hotel". There are three reviews listed:

- Review by Teresa**: We stayed one week at the Limerick Hotel and had a good experience there. The rooms are nice, clean and well decorated. Wi-Fi in the Lobby is free, but we wanted internet access in the room, so we paid an additional 15€ for the week.
- Review by Miguel**: A friend and I stayed here for a concert in Limerick city in November. We had to pay about 30 euros per night, which was very little. The room was really spacious, quiet and clean. Overall, great rooms for little money.
- Review by Curtis**: The hotel is located in the centre of Limerick and has a nice view of the river Shannon. But when I arrived in the room, the window was left open and it was freezing. All in all the hotel was ok, but I wouldn't book it again.

To the right of the reviews, there is a task instructions and three questions:

Lies die Bewertungen über das Limerick Hotel.

Du hast dir für jedes Hotel drei Fragen gestellt.

1. Beantworte die Frage mithilfe der Bewertungen. Klicke an!
2. In welcher Bewertung hast du gelesen, ob die Antwort ja oder nein ist?
Ziehe den Namen des Autors in das Kästchen.

Manchmal findest du in der gleichen Bewertung mehrere Antworten.

1. Gibt es im Zimmer gratis WLAN?
 ja nein
Review by ?
2. Ist das Hotel im Stadtzentrum?
 ja nein
Review by ?
3. Ist das Hotel preiswert (unter 35€ pro Nacht)?
 ja nein
Review by ?

Weiter

Reading: hotel search

IFB: "Trip to Limerick" (part 4)

The screenshot shows a web browser window with the following content:

- Address Bar:** Tools Limerick Comic Festival https://www.crazy4comics.com/Limerick-Comic-Festival/
- Page Title:** CRAZY4COMICS
- Navigation:** HOME | FORUM | CON INTEL | ABOUT US
- Section:** The Limerick Comic Festival: Sunday highlights by Nathan & Lucy
- Text:** Friday 8 p.m., Saturday 3 p.m. and Sunday 1 p.m.: Several film events will take place in Hall H. There will be anything from classic Science Fiction to short films. A detailed schedule can be found online and in the entry hall.
Sunday 9 a.m., Hall D: Fans of robotics will be in heaven in this discussion about the use of robots as toys and how our play will be changed in the future. You can also have a look at the latest inventions in robotic entertainment in Hall A. This exhibition will start on Saturday 11 a.m. and will be open for the rest of the day.
- Activity Area:** Höre den Podcast und lies den Blogeintrag.
Welche Veranstaltungen finden am Sonntag statt? Um welche Uhrzeit?
Ziehe die Kärtchen rechts zur passenden Uhrzeit.
Achtung:
Nicht alle 5 Veranstaltungen finden am Sonntag statt.
Du musst also nicht alle freien Kästen füllen.
Das können wir am Sonntag besuchen:
entweder oder
9:00 ?
10:00 ?
11:00 ?
11:30 ?
13:00 ?
Ausstellung: Neuheiten in der Robotik
Diskussion: Roboter - das Spielzeug der Zukunft
Filmfestival: Science-Fiction-Klassiker
Kinofestival: Superhelden
Kurzfilmfestival: Alte Cartoons in neuem Licht
Weiter

Reading and listening: planning activities

IFB: steps in R&D

	Scenarios	Component tests
Pre-piloting	Individual qualitative interviews	-
Piloting	Trial runs in classes Trial runs of stimulated recall and thinking aloud with individual students	Trial runs in small groups
Main study	Administration in class Thinking aloud and stimulated recall with individual students	Administration in class

Cognitive task validation in the IFB project

- Methods
 - Piloting, field study & main study: stimulated recall interviews and observation
 - Field study & main study: think-aloud protocols, stimulated recall interviews and user tracking
- Goal: collect evidence to improve the validity of the scenario-based tasks
 - Identify and correct problems in the tasks (usability, content, difficulty, wording)
 - *Investigate test taking processes and strategies used by students (strategic knowledge and strategic regulation)*
 - *Compare informational quality of think-aloud, stimulated recall and oral summaries of the input text*

Chapelle & Voss 2014; Cohen, 2006, 2014;
Kenyon & MacGregor 2012; Rossa, 2012

Contact

Peter Lenz

Institute of Multilingualism

e-mail: Peter.Lenz@unifr.ch

www.centre-multilingualism.ch