

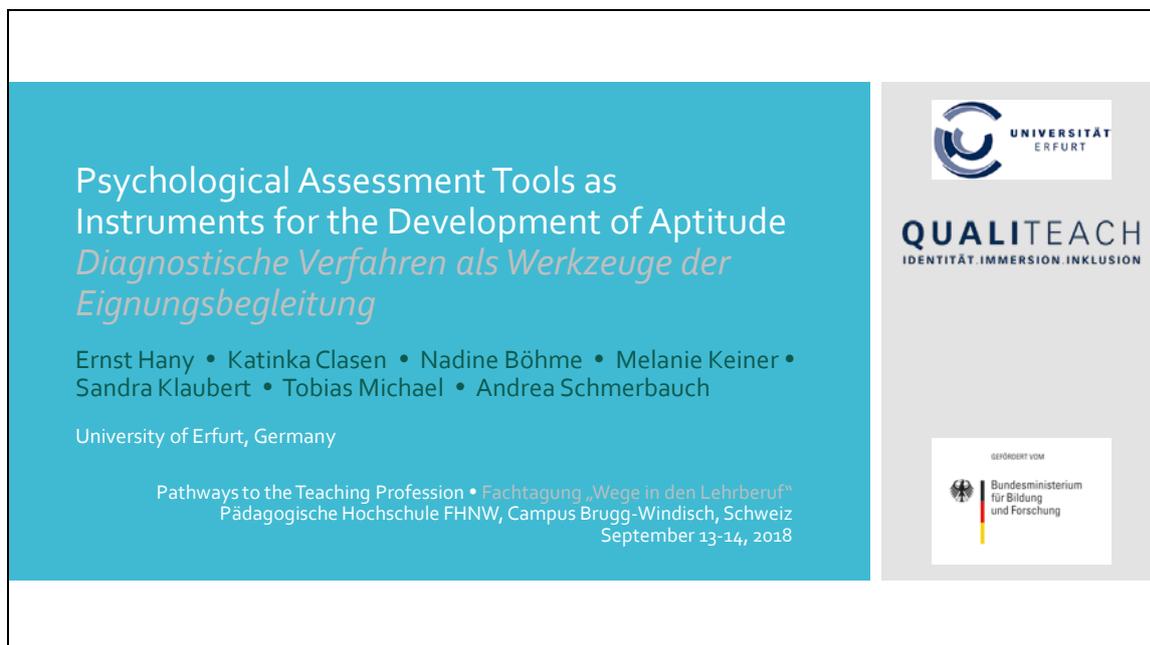
Psychological Assessment Tools as Instruments for the Development of Aptitude

Diagnostische Verfahren als Werkzeuge der Eignungsbegleitung

Ernst Hany • Katinka Clasen • Nadine Böhme • Melanie Keiner • Sandra Klaubert • Tobias Michael • Andrea Schmerbauch

University of Erfurt, Germany

Pathways to the Teaching Profession • Fachtagung „Wege in den Lehrberuf“
Pädagogische Hochschule FHNW, Campus Brugg-Windisch, Schweiz, September 13-14, 2018



The slide cover features a teal background on the left with white text. On the right, there is a white background with logos for the University of Erfurt, QUALITEACH, and the German Federal Ministry of Education and Research.

Psychological Assessment Tools as Instruments for the Development of Aptitude
Diagnostische Verfahren als Werkzeuge der Eignungsbegleitung

Ernst Hany • Katinka Clasen • Nadine Böhme • Melanie Keiner • Sandra Klaubert • Tobias Michael • Andrea Schmerbauch

University of Erfurt, Germany

Pathways to the Teaching Profession • Fachtagung „Wege in den Lehrberuf“
Pädagogische Hochschule FHNW, Campus Brugg-Windisch, Schweiz
September 13-14, 2018

UNIVERSITÄT ERFURT

QUALITEACH
IDENTITÄT. IMMERSION. INKLUSION

gefördert vom
Bundesministerium für Bildung und Forschung



The slide content has a teal header with the QUALITEACH logo and the title 'Teaching Talent Center'. Below the title is a group photo of seven people. To the right of the photo is a list of four bullet points.

QUALITEACH
IDENTITÄT. IMMERSION. INKLUSION

Teaching Talent Center

- In Germany, the federal government and the 16 states have invested funds into the programme „Quality Initiative for Teacher Education“.
- At the University of Erfurt, the project „QUALITEACH“ strives to develop the education for teachers (primary school, lower secondary school, special education).
- The „Teaching Talent Center“ is part of this project and offers counseling, coaching, and training to student teachers.
- Psychological assessment procedures provide the basis for the different forms of treatment.



Dear colleagues,

it is my pleasure and my honor to present at this wonderful conference. Our research is funded by the federal government under the heading “Quality Initiative for Teacher Education”. Many institutes of teacher education in Germany profit from these fundings during these years, and we have established a “Teaching Talent Center” which offers counseling, coaching, and training to teacher students. What we are presenting today, is an overview of some assessment instruments that we have developed and are currently using as a basis for counseling and training.

<h2>Table of Contents</h2>	<ul style="list-style-type: none">• Conceptual and strategic background• Appraisal of personality as a tool for improving self-knowledge and stimulating development<ul style="list-style-type: none">• Multi-dimensional questionnaire• Performance Assessment as a tool of developmental counseling<ul style="list-style-type: none">• Multiple Mini-Interviews
----------------------------	---

This is what you will hear from me in the next minutes: I will begin with a short introduction of the conceptual and strategic background of our work, followed by the presentation of a questionnaire for some personality traits, and after that, I will give you some information on the Multiple Mini-Interviews that we use as performance assessment tools.

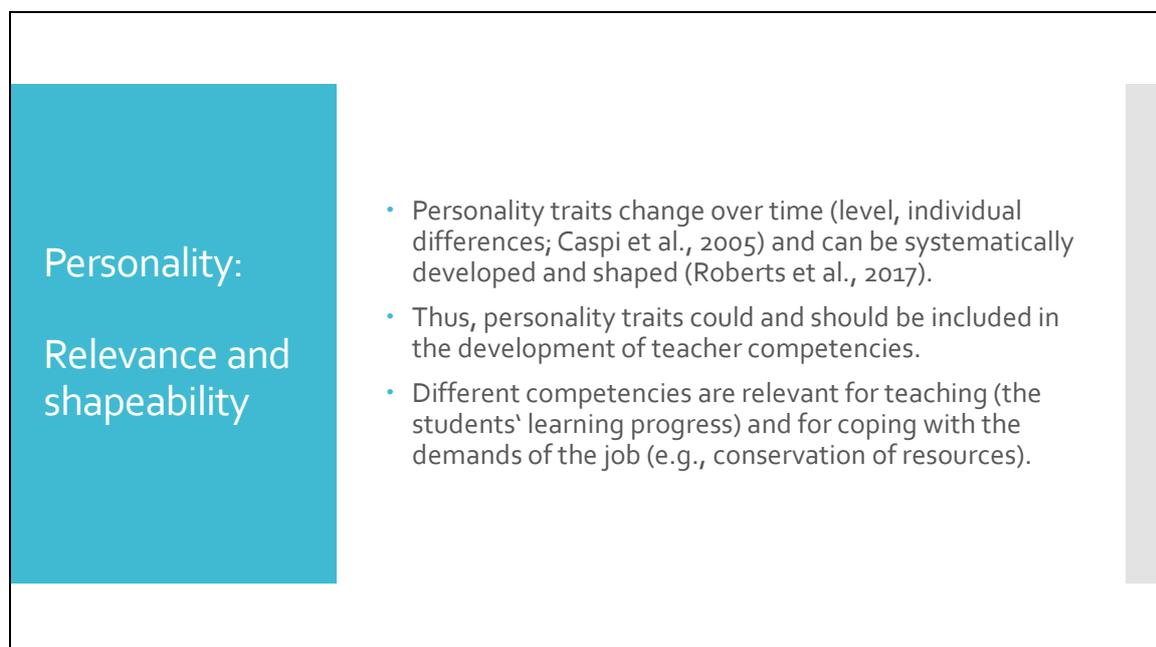
<h2>Personality: Relevance and shapeability</h2>	<ul style="list-style-type: none">• Personality has some relevance for occupational achievement (i.e., intelligence, Big Five, interests; Barrick et al., 2001).• Some personality traits (e.g, neuroticism, extraversion, conscientiousness) are also relevant for the teaching profession (Foerster, 2008; Nieskens, 2009).• Personality is more than the Big Five. It includes self-regulation, coping strategies, self-efficacy, or goal orientations, to name a few (McAdams & Pals, 2006).• Such traits are deeply connected to competencies which build the main paradigm for teacher education (Mayr & Neuweg, 2006; Weinert, 2001).
--	---

Why do we focus on personality? Personality has been shown to be rather stable through several periods of life, and it can be measured via questionnaires even at the beginning of the studies at the university – and this is what we do.

Personality has some relevance for occupational achievement in general. The correlations are not magically high but should not be neglected. Some personality traits have been demonstrated to be relevant for good teaching or the teachers' social behavior.

From the perspective of the social-cognitive theory of learning, personality consists not only of the Big Five but also of learned habits that include self-regulation, coping strategies, self-efficacy and more.

Such traits or trait-like behavioral tendencies are deeply connected to competencies and should be included in comprehensive models of the ingredients of competencies.



**Personality:
Relevance and
shapeability**

- Personality traits change over time (level, individual differences; Caspi et al., 2005) and can be systematically developed and shaped (Roberts et al., 2017).
- Thus, personality traits could and should be included in the development of teacher competencies.
- Different competencies are relevant for teaching (the students' learning progress) and for coping with the demands of the job (e.g., conservation of resources).

For our work, it is relevant that personality traits are of a certain stability but develop and can be changed – of course within certain limits. Many models of teacher competencies focus on the students' learning progress only (as criterion). We however focus also on the teachers' management of life, their work-life-balance, prevention of burn-out and other health-related aspects. And here, personality might be even more important than for the quality of teaching. With respect to teacher retention, the well-being and health of teachers' are of utmost importance.

Personality: Measurement

- A self-assessment questionnaire was designed to cover relevant personal qualities for the teaching profession.
- It includes six domains, with three to four subscales each (*mostly taken from existing instruments*):
 1. Interest in professional teaching activities
 2. Disdain of external incentives (e.g., long vacancies)
 3. Identification with the teaching profession
 4. Self-management considering personal resources
 5. Emotional stability
 6. Self-confident problem-solving
- Answers are collected through five-point Likert scales.

Thus, we have designed a questionnaire that covers several relevant personal qualities for the teaching profession. For easy communication, we have restrained our focus on six domains. Each domain is assessed through three to four subscales, and most of these scales are taken from existing instruments because we didn't want to reinvent the wheel.

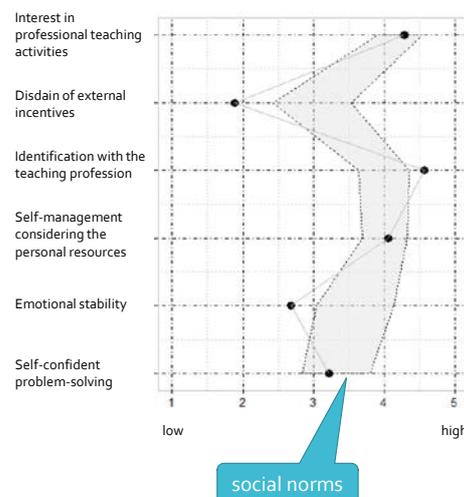
The six personality domains read as follows:

1. Interest in professional teaching activities (like preparing lessons and giving instruction)
2. Disdain of external incentives (that is, the disregard of extrinsic features of the job, like long vacancies or having a secure job – because preferring this might be undermine study motivation and satisfaction with the real job)
3. Identification with the teaching profession
4. Self-management considering personal resources (that is, preserving as good work-life-balance, including being able to find an emotional distance to job issues)
5. Emotional stability
6. Self-confident problem-solving

Personality: Feedback

- Written feedback is given comprising several pages.
- Norm- and criterion-referenced feedback is provided.
- The meaning of the scales and its extremes are described in an information booklet.
- Personal feedback is offered in small groups.
- Counseling sessions are based on the results.

Example of a visual feedback sheet

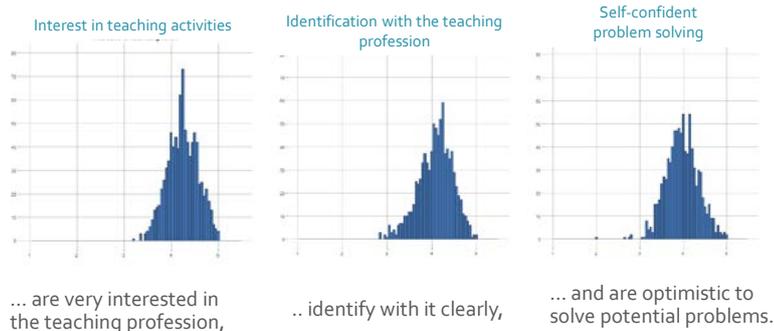


We offer our questionnaire to our students at the beginning of their academic life, and provide them with written or personal feedback, as they wish. The feedback is based on some graphical displays. You see an example of the main feedback graphic comprising the six major domains. The participants are provided with their individual results. These are framed within social norms (the shaded areas which include the scores of 70 percent of their peers) and within criterion-related norms (this is, the original scale values).

Feedback is offered in small groups but individual counseling sessions can be arranged too.

Personality: Empirical distribution of scale values

Most of the students (total n = 877) ...

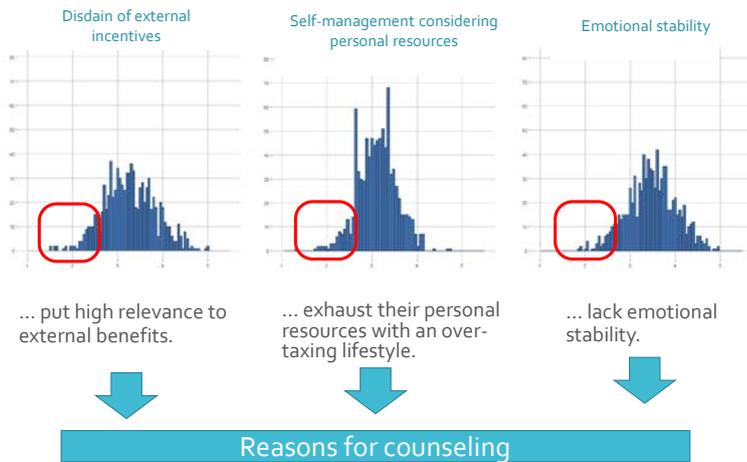


There is an overall positive trend (high means) but also remarkable individual differences.

If we take a short look at the results, we find that the teacher students generally reach positive scores in the domains of teaching interests, identification, and problem-solving. That means, that the general tendency is positive but there is also a large variance between the students.

Personality:
Empirical distribution of scale values (cont.)

Some teacher students ...



If we turn to the more problematic domains (it's not the "dark triad" of personality but something similar), we find not so positive mean values and even larger distributions of individual scores. There are substantial groups of students who value highly the extrinsic incentives to the teaching job, who are not very skilled in managing life stress, and who lack emotional stability.

When we see such scores on the individual profiles, we recommend specific training courses to the students, for example, to improve time management or conflict communication.

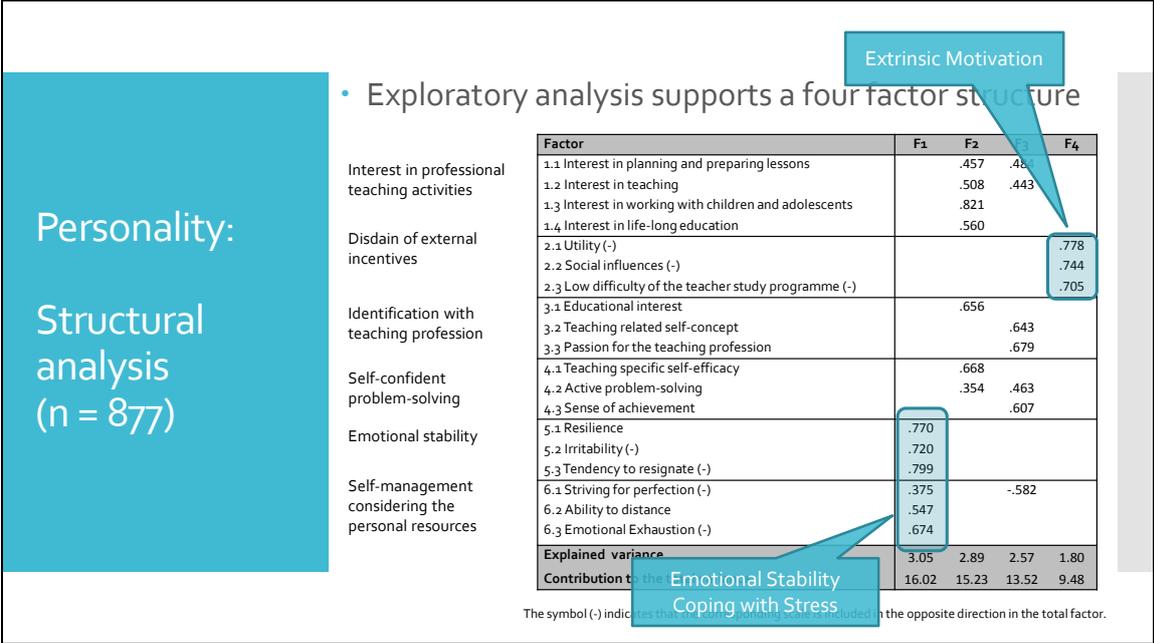
Personality:
Structural analysis (n = 877)

• Exploratory analysis supports a four factor structure

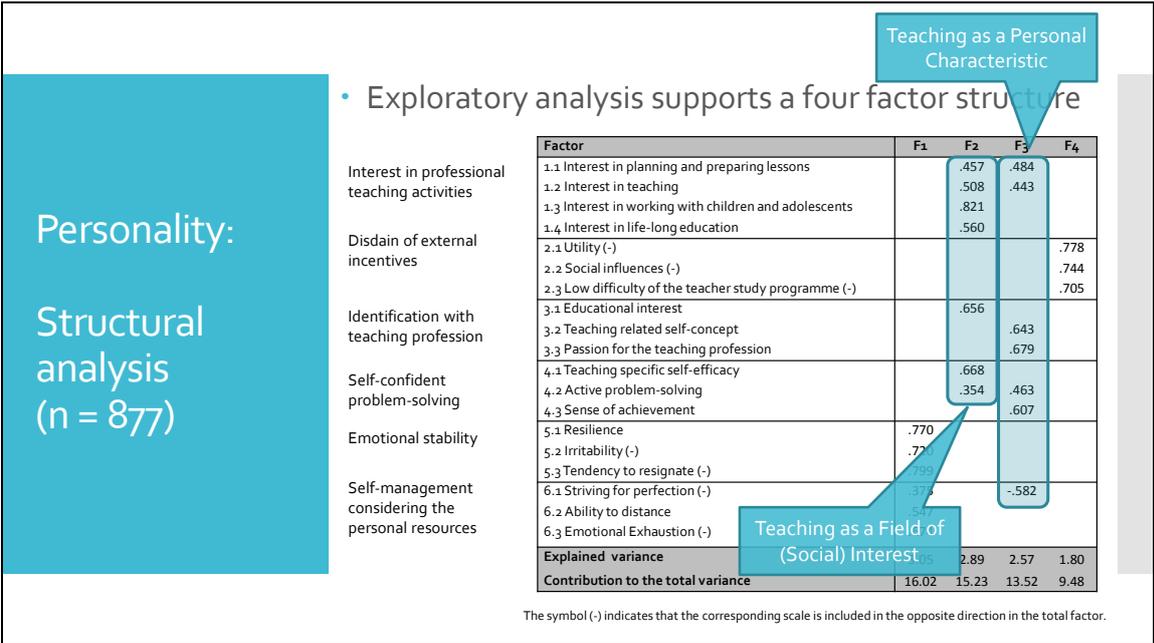
Factor	F ₁	F ₂	F ₃	F ₄
Interest in professional teaching activities				
1.1 Interest in planning and preparing lessons		.457	.484	
1.2 Interest in teaching		.508	.443	
1.3 Interest in working with children and adolescents		.821		
1.4 Interest in life-long education		.560		
Disdain of external incentives				
2.1 Utility (-)				.778
2.2 Social influences (-)				.744
2.3 Low difficulty of the teacher study programme (-)				.705
Identification with teaching profession				
3.1 Educational interest		.656		
3.2 Teaching related self-concept			.643	
3.3 Passion for the teaching profession			.679	
Self-confident problem-solving				
4.1 Teaching specific self-efficacy		.668		
4.2 Active problem-solving		.354	.463	
4.3 Sense of achievement			.607	
Emotional stability				
5.1 Resilience	.770			
5.2 Irritability (-)	.720			
5.3 Tendency to resignate (-)	.799			
Self-management considering the personal resources				
6.1 Striving for perfection (-)	.375		-.582	
6.2 Ability to distance	.547			
6.3 Emotional Exhaustion (-)	.674			
Explained variance	3.05	2.89	2.57	1.80
Contribution to the total variance	16.02	15.23	13.52	9.48

The symbol (-) indicates that the corresponding scale is included in the opposite direction in the total factor.

An exploratory factor analysis of our 19 scales did not confirm our six domains as independent facets of the teacher personality. By the way, we've got the same result when we used a confirmatory analysis. Instead, four independent factors emerged.



The first factor represents Emotional Stability and Coping with Stress, and the fourth factor equals our Extrinsic Motivation Domain.



The Factors 3 and 4 divide our scales into a kind of social interest dimension and into a kind of self-concept dimension. It would be interesting to follow this distinction further. Maybe we find students who choose the teaching profession in a more rational way, and others who choose it in more personal and emotional way. However, our factors are unrelated, and that means, that all combinations of individual factor values are possible.

Personality

Relationships between personality and occupational choice indicators (n = 877)

	Personality domains					
	Interest in teaching	Disdain of external incentives	Identification	Self-confident problem-solving	Emotional stability	Self-management
Personality domains						
Interest in teaching	-					
Disdain of external incentives	.08*	-				
Identification	.54**	-.03	-			
Self-confident problem-solving	.52**	.03	.55**	-		
Emotional stability	.32**	.19**	.31**	.45**	-	
Self-management	.02	.12**	-.05	.07*	.45**	-
Occupational choice indicators						
In-Breadth Career Exploration	.05	-.14**	-.02	.02	-.04	-.08*
In-Depth Career Exploration	.41**	-.05	.49**	.49**	.22**	-.04
Career Commitment	.36**	.02	.49**	.42**	.27**	.04
Identification with Career Commitment	.40**	-.08*	.55**	.49**	.23**	.01
Career Commitment Flexibility	-.34**	-.16**	-.40**	-.37**	-.34**	-.10**
Career Self-Doubt	-.33**	-.11**	-.37**	-.43**	-.40**	-.20**

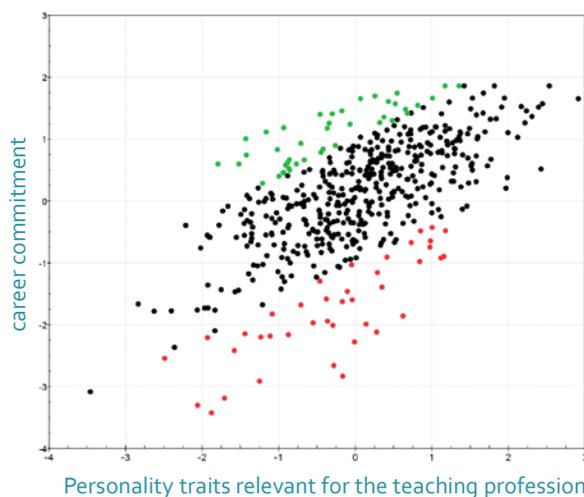
* correlation is significant at the level of $p < .05$; ** correlation is significant at the level of $p < .01$

Since the six fields of personality are not factorially separated, the major scales are correlated as we see in the upper part of this table. However, we are willing to continue to refer to the six domains because they are easy to communicate and the students easily understand what they mean.

In terms of criterion validity, we find substantial correlations with indicators of career choice. These are the scales developed by Porfeli and colleagues to assess the Career Identity Status. That means, that the self-attribution of characteristics that are relevant for the job as a teacher corresponds to the strength of the career choice with respect to becoming a teacher.

Personality

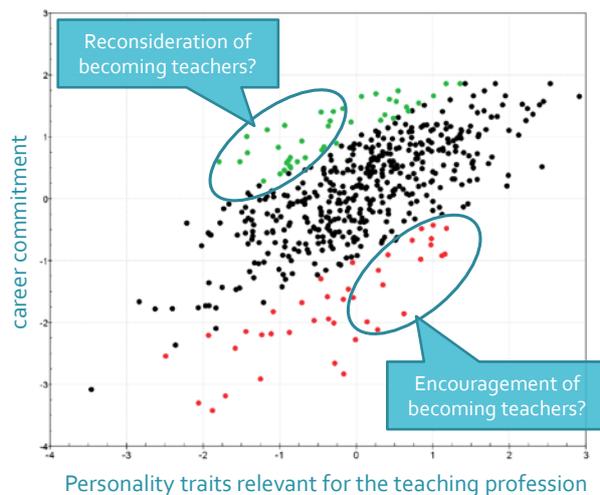
Relationships between personality and occupational choice indicators (n = 877)



However, this relationship is far from perfect. We have summed up some personality traits and some scales for the career commitment and find a close but a not too close relationship.

Personality

Relationships between personality and occupational choice indicators (n = 877)



And this gives us reasons for specific counseling approaches. Those students who are overcommitted to the teaching profession should be motivated to reflect on their not so convincing aptitude. And those students who are not so much committed to the teaching profession despite proper trait values, could be encouraged to become teachers actually.

Performance Assessment:

How Multiple Mini-Interviews (MMIs) work

- Multiple Mini-Interviews are intended to replace long-lasting assessment procedures involving a number of observers.
- Each interview presents a unique problem situation that has to be solved verbally or manually within a few minutes.
- Each situation is intended to independently test one relevant competence for the aspired occupation.
- Behavioral performance is often evaluated by a global rating scale only.



Let me turn to our approach for the Performance Assessment. Not all characteristics of good teachers can be assessed by self-description. We need some standardized procedures to observe what they really do and can do.

Now, the method of Multiple Mini-Interviews consists of several short performance situations like in an assessment center. A defined problem or question is presented to the participants and has to be solved or answered within a few minutes.

Performance Assessment:

How Multiple Mini-Interviews (MMIs) work

- MMIs have been designed at McMaster University (Canada) to be used in Medical School admissions (Eva et al., 2004).
- MMIs are nowadays widely used and accepted as student selection method in the health professions (Pau et al., 2013).
- MMI scores correlate substantially with final practical exam performance (Pau et al., 2013).
- They correlate low with traditional aptitude indicators (Knorr & Hissbach, 2014) which constitutes their incremental validity.

The Multiple Mini-Interviews have been designed in Canada as a method for admission to the Medical School. From that, the method has spread to other institutions and other study programs. Research has shown that MMI scores collected at the beginning of the study program correlate substantially with the final performance in practical exams. The MMI scores correlate low with traditional aptitude indicators. Thus, both approaches to measure aptitude should complement each other.

Performance Assessment:

How reliable are ratings in Multiple Mini-Interviews?

- One rater's reliability of lesson quality is about .37, and only by using four different raters judging four different lessons, reliability increases to .72 (Ho & Kane, 2013).
- While classroom management can be evaluated with reliability $>.80$ through two observed lessons only, reliable evaluation of learning support requires about 20 observed lessons (Praetorius et al., 2014).
- MMI rating reliability is generally described as satisfactory to good (Pau et al., 2013; Rees et al., 2016).
- Correlations between two raters at the same MMI station (rating $n=444$ subjects) range from .41 to .69 (Sebok et al., 2014).

Now, there is a big problem with the judgment of behavior, that is, the lack of reliability that has been found in a lot of studies. If behavior is judged by human raters, the judgments often represent mainly the individual standards of the raters and less the quality of the demonstrated behavior.

With respect to the MMI method, the results regarding getting reliable scores are mostly positive but not too positive.

Performance Assessment:

Research question

- **How reliable are MMI ratings of teacher students' behavior?**
- Material
 - Rubrics for six MMI stations with several indicators and four (ordinal) categories for each indicator
 - Up to ten raters for each recorded situation
- Sample
 - 99 students (mean age 23.0 years, $sd = 2.30$; 89.9 % female), completing six MMI stations
- Reliability score
 - Calculation of intraclass correlation coefficients (ICC)
 - ICC(2,1) reports the reliability for one randomly selected rater
 - ICC(2,k) reports the reliability for the mean of k raters

Thus, we were aware that we had to secure the reliability of the judgments before we could use the scores for individual feedback.

And this is what we did: We developed six MMI stations with performance tasks and rubrics to assess the behavior shown. The rubric for each station consisted of five to eight indicators and for each indicator, of four ordinal stages of quality.

99 students completed these six MMIs. All behaviors were recorded, and a sample of ten recordings was evaluated by up to ten raters.

To calculate an index of reliability, we used the intraclass correlation coefficient. This statistic calculates the proportion of variance of the judgments that is related to the differences between the performers.

Example of a MMI task

MMI task

„Dealing with an awkward caretaker at school“

- Candidates are handed out a brief task sheet
- They are given 2 minutes preparation time

“For your English lesson today, you have announced a listening test. Thus you brought your own CD player but unfortunately – as you have just noticed in the faculty room – it does not work. A colleague told you that you might borrow a device from Mr Mueller, the official caretaker at your school. You have to hurry because your lesson will start in 5 minutes and students must not be left alone. You arrive at the caretaker's door.”

Now I want to introduce to you an example of the MMI tasks we used. This is an interactive task that requires communication with a real person. The participants got the instruction that they

should imagine to be at school and that they were in need of a compact disk player. They should ask the caretaker of the school to hand them out such a device. Because this situation was framed to happen within the recess time, the participants had five minutes only to get the device.

Example



- 5 minutes interaction with the awkward caretaker who worked friendly but very slowly (an instructed student actor).
- The interactions were video-recorded.
- Performance was evaluated on the basis of five rubrics.
- Nine experienced raters evaluated ten performances, to get a basis for the ICCs.

The problem was that the caretaker was very friendly but also very slowly and the participant had to take several means to speed up his actions while staying polite and supportive.

All performances were recorded and evaluated. Ten performances were then rated by nine team members.

ICCs for the calibration sample

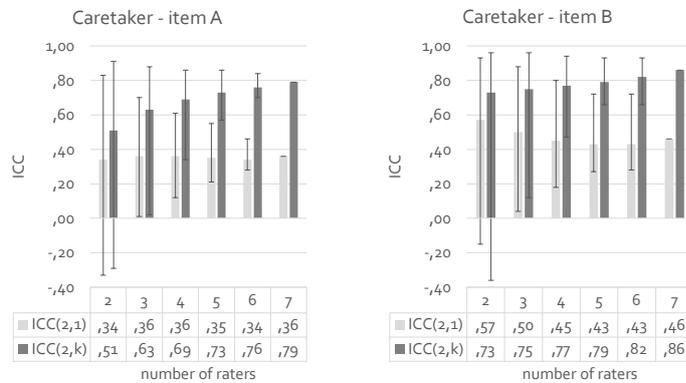
Performance indicators	ICC(2,1)	ICC(2,9)
Taking initiative, hurrying the caretaker	.43	.87
Actively supporting the caretaker's work	.47	.89
Emphasizing pressure of time	-.01	-.06
Politeness (affective tone)	.62	.94
Relaxed but focused appearance	.22	.72

Each ICC represents the proportion of the variance of the ratings which is attributable to the „true differences“ of the subjects.
 ICC(2,1) focuses on one average rater, ICC(2,9) assumes nine active raters.

This table shows the ICC values of the five performance indicators for this situation. As you might recognize, three of the five indicators had a good reliability, provided that they were judged by nine raters. However, if only one rater was available, none of the reliability coefficients would have turned out to be satisfying.

Results:
Interrater reliability

Intraclass correlation: How many raters are needed for reliable judgements?



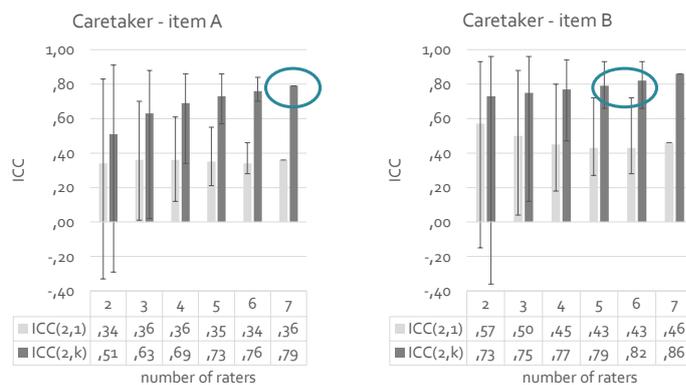
Median, minimum and maximum ICCs are shown.

Now we want to take a closer look at the reliability scores and we focus on just two indicators.

These graphics show the ICCs obtained by two to seven raters, given that all possible combinations of n raters out of the nine raters were used to calculate the scores. What you see are summarizing statistics of the ICCs scores that were calculated on the basis of all possible subgroups of raters.

Results:
Interrater reliability

Intraclass correlation: How many raters are needed for reliable judgements?

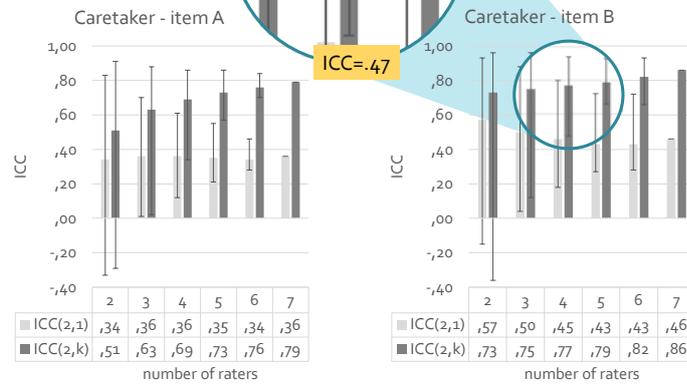


Median, minimum and maximum ICCs are shown.

What we see first is that the items are not uniformly apted to provide a reliable rating. For item A, seven raters are necessary to obtain a coefficient of about .80 whereas for item B, only five to six raters are needed. Thus the items to be rated must be designed carefully to produce uniform ratings.

Results: Interrater reliability

Intraclass correlation coefficients for item B are needed for reliable results



Second, we see that different groups of raters can produce very different ICC values. Here you see a magnified detail of the graphics for item B. Having four raters available, they can produce ICCs between .47 and .94. So the reliability of the results depends not only on the number of the raters but also on the similarity of their rating standards. It is therefore important to carefully select or calibrate the raters if you want to obtain reliable results.

Measures to improve the results

- The categories of the rubrics were more clearly defined and enriched with adequate examples.
- The final raters received a detailed training.
- The final raters assessed the video recordings in packages of 25 cases each, and then reviewed their ratings together to adjust their individual perspectives.

Now, how did we deal with these problems? First, we redesigned the categories of the rubrics to make them more precise. Second, we trained the raters with several elaborated cases. Third, the raters of the 99 performances reviewed their ratings together, each time after 25 performances had been rated. That helped to align their standards.

Caretaker example:

ICCs for the final *two raters* and *all 99 cases*

Performance indicator	ICC(2,1)	ICC(2,2)
Taking initiative, hurrying the caretaker	.75	.85
Actively supporting the caretaker's work	.69	.81
Emphasizing pressure of time	.76	.87
Politeness (affective tone)	.74	.85
Relaxed but focused appearance	.54	.70

Each ICC represents the proportion of the variance of the ratings which is attributable to the „true differences“ of the subjects.
 ICC(2,1) focuses on one average rater, ICC(2,2) assumes two active raters.

After having rated all 99 performance, we calculated the ICC values again and got very good scores, provided that two well trained raters judge every performance.

Results:
Interrater reliability

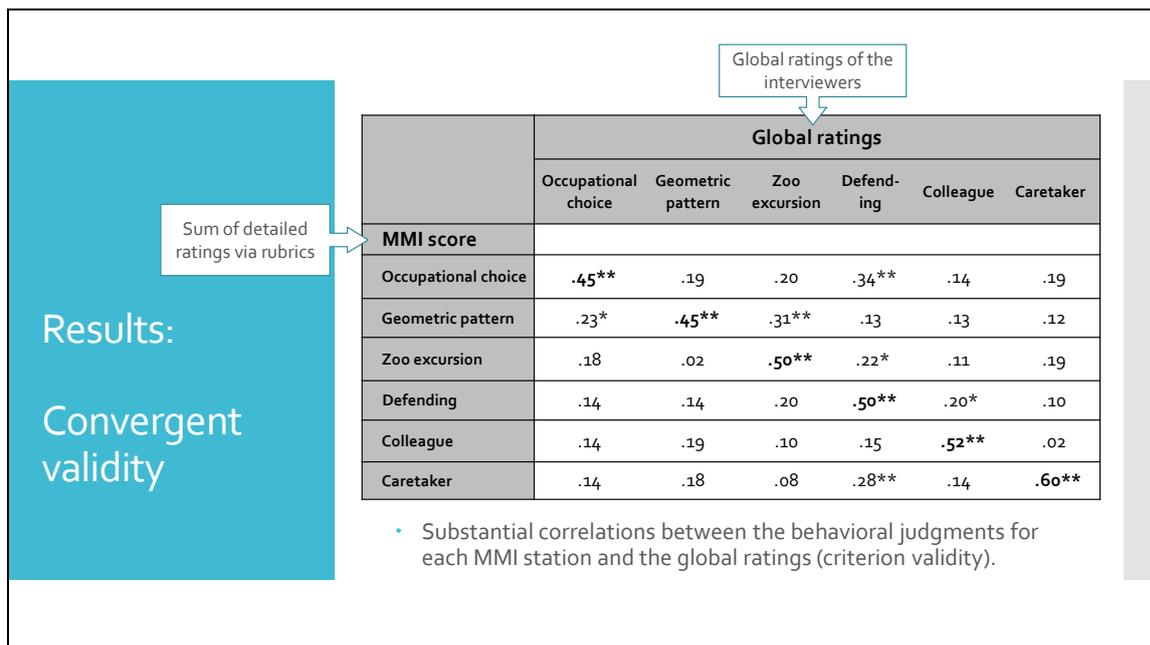
Intraclass correlations (two raters per situation)

- The ranges refer to the criteria to be judged in each situation -

MMI station	ICC(2,1)	ICC(2,2)
Description of a geometric pattern	.61 - .90	.75 - .95
Justifying the occupational choice	.15 - .86	.26 - .92
Planning a zoo excursion	.34 - .72	.51 - .83
Defending the teaching job against provocative criticism	.46 - .83	.63 - .91
Discussion with a difficult colleague	.52 - .74	.66 - .85
Handling the awkward caretaker	.54 - .76	.70 - .87

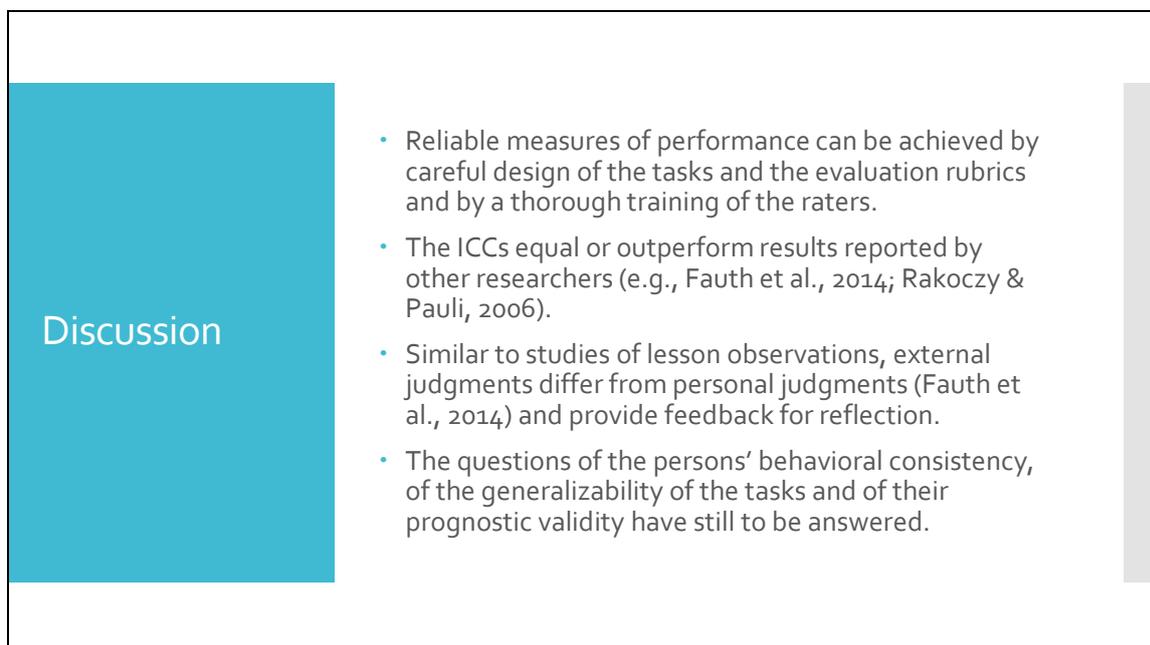
- In each MMI situation, some items can be rated with high reliability.
- Some items fail to reach sufficient reliability.
- Two raters are more reliable than one rater.

We did so with each of our six MMI situations. This table shows the ICC ranges for the performance indicators for each situation. In each situation, we have some very reliable indicators but there are also some indicators that cannot easily be rated in a concordant manner. This means more work to be done.



To give you a tiny impression of the validity tests, I show you a table with correlation coefficients. To calculate these coefficients, the scores based on the rubrics were summed up and compared to the global ratings provided by the staff who supervised the recordings.

The correlations show a consistent pattern of convergent and divergent validity. Their moderate level might justify that we use detailed rubrics instead of global ratings.



We conclude from our results that it is possible to produce reliable performance assessments. However, this needs considerable effort. The ICCs that we obtain are comparable to other studies. This encourages us to continue our work.

However, reliability is just one criterion of quality of assessments. The questions of behavioral stability, of the generalizability of the tasks (or their representativeness), and of their prognostic validity have still to be investigated. It's always the same: "More research is needed".

Thank you for
your attention!



Reports and Material
(in German language)

References

- Abdul Rahin, A. F. & Yusoff, M. S. B. (2016). Validity evidence of a Multiple Mini Interview for selection of medical students: Universiti Sains Malaysia Experience. *Education in Medicine Journal*, 8(2), 49-63.
- Barrick, M., Mount, M., & Judge, T. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, 9, 9-30.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9 (4), 469-520.
- Berliner, D. C. (1992). The nature of expertise in teaching. In F. K. Oser, A. Dick & J.-L. Patry (eds.), *Effective and responsible teaching – the new synthesis* (pp. 227-248). San Francisco: Jossey-Bass.
- Brown, A. & Bimrose, J. (2015). Identity development. In P. J. Hartung, M. L. Savickas & W. B. Walsh (Eds.), *APA handbook of career intervention. Volume 2: Applications* (pp. 241-254). Washington, CD: APA.
- Caspi, A., Roberts, B. W., & Shiner, R. L. (2005). Personality development: Stability and change. *Annual Review of Psychology*, 56, 453-484.
- Dawis, R. V., England, G. W., & Lofquist, L. H. (1964). *A theory of work adjustment*. Minnesota Studies in Vocational Rehabilitation (No. XV), 1-27. Minneapolis: University of Minnesota, Industrial Relations Center.
- Erpenbeck, J. & Rosenstiel, L. v. (2007). *Handbuch Kompetenzmessung. Erkennen, verstehen und bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis* (2nd ed.). Stuttgart: Schäffer-Poeschel.
- Eva, K., Rosenfeld, J., Reiter, H. & Norman, G. (2004). An admissions OSCE: The multiple mini-interview. *Medical Education*, 38(3), 314-326.
- Foerster, F. (2008). *Personale Voraussetzungen von Grundschullehrerstudierenden. Eine Untersuchung zur prognostischen Relevanz von Persönlichkeitsmerkmalen für den Studien- und Berufserfolg*. Münster: Waxmann.

References (cont.)

- Frey, A. (2008). Kompetenzstrukturen von Studierenden in der ersten und zweiten Phase der Lehrerbildung. Eine nationale und internationale Standortbestimmung. Landau: Verlag Empirische Pädagogik.
- Gafni, N., Moshinsky, A., Eisenberg, O., Zeigler, D. & Ziv, A. (2012). Reliability estimates: Behavioural stations and questionnaires in medical school admissions. *Medical Education*, 46(3), 277-288.
- Holland, J. L. (1997). Making vocational choices: A theory of vocational personalities and work environments (3rd ed.). Odessa, FL: Psychological Assessment Resources.
- Knorr, M. & Hissbach, J. (2014). Multiple mini-interviews: same concept, different approaches. *Medical Education*, 48(12), 1157-1175.
- Mayr, J. (2009). LehrerIn werden in Österreich: empirische Befunde zum Lehramtsstudium. *Erziehung und Unterricht*, 159/1-2, 14-33.
- Mayr, J., & Neuweg, G. H. (2006). Der Persönlichkeitsansatz in der Lehrer/innen/forschung. Grundsätzliche Überlegungen, exemplarische Befunde und Implikationen für die Lehrer/innen/bildung. In M. Heinrich & U. Greiner (eds.), *Schauen, was rauskommt. Kompetenzförderung, Evaluation und Systemsteuerung im Bildungswesen* (pp. 183-206). Wien: Lit.
- Meadams, D. P., & Pals, J. L. (2006). A new big five fundamental principles for an integrative science of personality. *American Psychologist*, 61(3), 204-217.
- Nieskens, B. (2009). Wer interessiert sich für den Lehrerberuf – und wer nicht? Berufswahl im Spannungsfeld von subjektiver und objektiver Passung. Göttingen: Cuvillier.
- Pau, A., Jeevaratnam, K., Chen, Y. S., Fall, A. A., Khoo, C. & Nadarajah, V. D. (2013). The multiple mini-interview (MMI) for student selection in health professions training – A systematic review. *Medical Teacher*, 35, 1027-1041.
- Pawlik, K. (1982). Modell- und Praxisdimensionen psychologischer Diagnostik. In K. Pawlik (ed.), *Diagnose der Diagnostik* (2nd ed., pp. 13-43). Stuttgart: Klett.

References (cont.)

- Rees, E. L., Hawarden, A. W., Dent, G., Hays, R., Bates, J. & Hassell, A. B. (2016). Evidence regarding the utility of multiple mini-interview (MMI) for selection to undergraduate health programs: A BEME systematic review: BEME Guide No. 37. *Medical Teacher*, 38, 443-455.
- Roberts, B. W., Luo, J., Briley, D. A., Chow, P. I., Su, R., & Hill, P. L. (2017). A systematic review of personality trait change through intervention. *Psychological Bulletin*, 143(2), 117-141.
- Sebok, S. S., Luu, K. & Klinger, D. A. (2014). Psychometric properties of the multiple mini-interview used for medical admissions: Findings from generalizability and Rasch analyses. *Advances in Health Sciences Education*, 19(1), 71-84.
- Shulman, L. S. (1987). Knowledge and teaching: Foundation of the new reform. *Harvard Educational Review*, 57 (1), 1-22.
- Weinert, F. E. (2001). Concept of competence. A conceptual clarification. In D. S. Rychen & L. H. Salganik (eds.), *Defining and selecting key competencies* (pp. 45-65). Göttingen: Hogrefe.
- Zytowski, D. G. (2015). Test interpretation: Talking with people about their test results. In P. J. Hartung, M. L. Savickas & W. B. Walsh (Eds.), *APA handbook of career intervention. Volume 2: Applications* (pp. 3-10). Washington, CD: APA.