Do teacher students' competences develop in practical field experiences and does university training support this development?

Longitudinal findings from the KOSTA project

Dr. Christoph Schneider, Dr. Rainer Bodensohn
University of Koblenz-Landau, Campus Landau, Germany

KOSTA: Survey of Students in Teacher Education
at Landau University (Germany)

- Survey of <u>all</u> students in pre-service teacher education at the university

- Longitudinal evaluation of the competence development in the course of the studies (KOSTA project)

- Start of survey in 2007

    o t0: begin of studies (upon entering university)

    o t1: at the end of a compulsory three-week-long practical field experience; usually after the 3rd or 4th semester; *before deciding on a school type*

    o t2: at the end of another compulsory three-week-long practical field experience; usually at the end of the Bachelor of Education studies; *Time interval between t1 and t2: 1 to 1½ years*

The KOSTA Sample

N = 429 cases with full longitudinal data

- 22.9% males

- Age upon entering university 21.2 (± 3.8) years; 10% older than 25 years

- 17.2% had begun another university study programme beforehand;
  15.5% hat completed full vocational training in another field

- Distribution of school types (t2)

  - Primary school (up to grade 4):   42.7%

  - Lower-track secondary school (grades 5 to 10):   19.1%

  - Special needs education: 20.5%

  - Higher-track secondary school (grades 5 to 13):   16.1%

  - undecided / other: 1.6 %

- => Sample can *not* claim overall representativity for the "students in teacher education" population, but does comprise all careers

Output Orientation in Teacher Education(?)

- Strongly advocated by Cochran-Smith (2001)

**From the ETUCE(2008) policy paper
on *Teacher Education in Europe*:**
" ...the emphasis is not so much on the actual contents of the curricula but on the competences that teachers are able to show throughout their career. There is a shift in emphasis from certifica-tion based on the mastery of curricula to certification based on competences.
[....But t]he competences described must not be low-level, narrowly defined behaviourist specifications of knowledge and actions; rather they should be high-level, broadly defined statements of the characteristics which mark teachers"
A German saying goes: "wash me, but don't make me wet"

Contents of Surveys
„Standards for Teacher Training; Educational Sciences" (KMK, 2004; 2011): Eleven competences* in four domains

- ## Competence domain CD1 „Teaching"

  - o C1: "Teachers plan their lessons properly and behave professionally in the classroom"

  - o C2: "Teachers support student learning by providing adequate learning arrangements; they motivate students and encourage them to use the learning matters for transfer purposes"

  - o C3: "Teachers promote students' self-regulated learning and working"

- ## Competence domain CD2 „Education"

  - o C4: "Teachers are familiar with their students' social and cultural living conditions; in the school context, teachers influence students' individual development"

  - o C5: "Teachers impart values and norms; they support students' self-directed judging and acting"

  - o C6: "Teachers find pathways for resolving difficulties and conflicts in school and in the classroom"

  *NOTE: Competence descriptions are non-official translations for illustrative purposes

Contents of Surveys
„Standards for Teacher Training: Educational Sciences"
(KMK, 2004; 2011): Eleven competences in four domains

- Competence domain CD3 **„Assessment"**

    o C7: „Teachers diagnose students' prerequisites for learning and learning processes; they support individual students and counsel learners and their parents"

    o C8: "Teachers assess students' learning outcomes by means of transparent evaluation criteria"

- Competence domain CD4 **„Innovation"**

    o C9: "Teachers are well aware of their professions' particular demands; they understand their profession as a public appointment with a high level of responsibility and commitment"

    o C10: "Teachers understand their profession as a perpetual learning task"

    o C11: „Teachers contribute to the planning and the implementation of school development projects and programmes"

Operationalization of the KMK's Standards within the KOSTA Framework

- For each of the eleven competences, experts formulated five to eight items on the basis of a qualitative literature review (total: 65 items)

- Item wordings express behaviour elements ("competence elements") in accordance with the standards, these elements are meant to observable in the classroom or in school

- For each item, three aspects were rated:

  a) **Rating of the standard's overall importance**

  b) **Frequency of application [in proper teaching endeavours]**

  c) **Perceived quality of university training [towards the standard]**

KOSTA Item Example

- Students' self ratings referring to each competence element (item) referring to three aspects

5.	I flexibly deal with learning or comprehension problems. *(from C1)*		
a	Which overall importance do I assign to this competence?	(1) very important ... (6) completely unimportant	1 2 3 4 5 6
b	How often did I apply this competence?	(1) very often ... (6) never	1 2 3 4 5 6
c	How well did the university train me towards this competence?	(1) very well ... (6) not at all	1 2 3 4 5 6

Preliminary Studies on Structural Aspects
of the Measures Used

1. Do the item sets (of competence elements) assigned to each of the eleven competences constitute reliable units of observation?

   o *Method: Reliability analysis*

   o Summary: On a whole, the observed competence elements seem to form reasonably reliable constructs in the sense of the normatively proposed structure

2. Is it possible to empirically distinguish the four broad competence domains postulated by the KMK?

   o *Method: Confirmatory factor analysis (CFA)*

   o Summary for II: The normative assumption that the eleven competences are indicative of four broader competence domains is widely backed. It seems appropriate to model these four competence domains by the eleven competences as indicators for the purpose of further analysis
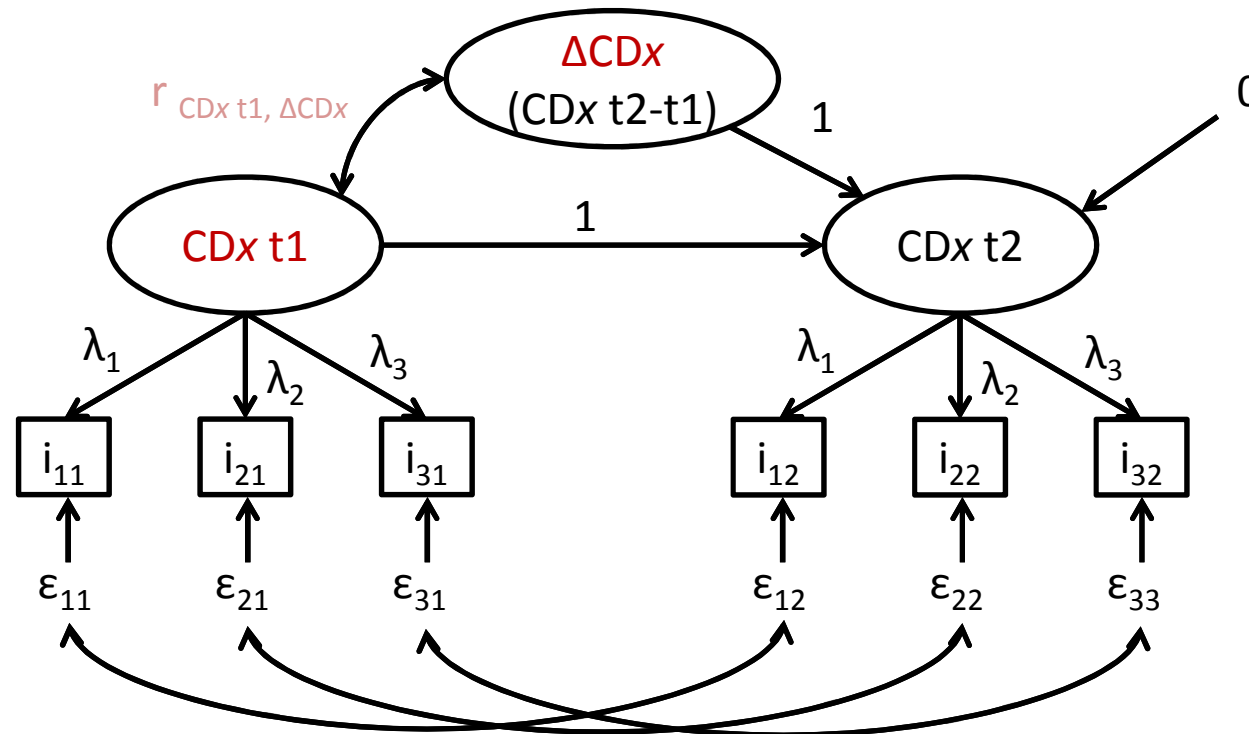
Main Research Questions

I.   Is there a (positive) longitudinal development of competences within each domain…?

   o *Method: Latent change modelling (LCM)*

II. … and does the perceived quality of university training support this development?

   o *Method: Latent variable modelling*

I. Is there a (positive) longitudinal development of competences within each domain…?

I. Is there a (positive) longitudinal development of competences within each domain…?

o   Latent change model structure (Geiser, 2010) encompassing the modelling of indicator *correlated uniqueness*:

I. Is there a (positive) longitudinal development of competences within each domain…?

a. Fit of the latent change models with mean structures

„Frequency"	X²	df	RMSEA	NNFI	CFI	sRMR
CD1 „Teaching"	21.9	9	.058	.982	.989	.076
CD2 „Education"	14.3	9	.037	.994	.997	.034
CD3 „Assessment"*	20.8	2	.148	.929	.976	.054
CD4 „Innovation"	28.2	9	.070	.968	.981	.064

- CD1 model: acceptable fit

- CD2 model: good fit

- CD3 model: poor fit, obvious problems, hardly suitable for interpretation

- CD4 model: (borderline) acceptable fit

* CD3 has only two indicators; here, a latent methods factor (M-1) was introduced instead of *correlated uniqueness*

I. Is there a (positive) longitudinal development of competences within each domain…?

b. Latent parameter values (Estimates) and lat. correlation of initial parameter values and difference scores

„Frequency"	M (CDx t1)	Var (CDx t1)	M (ΔCDx)	Var (ΔCDx)	$r_{CDx\ t1,\ \Delta CDx}$
CD1 „Teaching"	**4.67**	.19 ***	**.19 ***	.15 ***	-.59 ***
CD2 „Education"	**4.18**	.48 ***	**.07 ns**	.42 ***	-.53 ***
CD3 „Assessment"	3.43	.82 ***	.21 ***	.62 ***	-.52 ***
CD4 „Innovation"	**4.46**	.25 ***	**.21 *****	.25 ***	-.45 ***

1.) Absolutely seen, the parameter values are fairly high already at t1 (from 3.43 to 4.67, referring to the range of 1 to 6)

2.) In CD1 and CD4, there is a significant increase between t1 and t2; numerically, this increase is relatively low (.19 to .21; roughly 2/5 SD units)

3.) Across all CDs, there is a high negative correlation between initial (t1) level and difference score

II … and does the perceived quality of university training support this development?

II. … and does the perceived quality of university training support this development?

- Modelling the impact of "Perceived Quality Training" (PQoT) on "Frequency" (Freq):
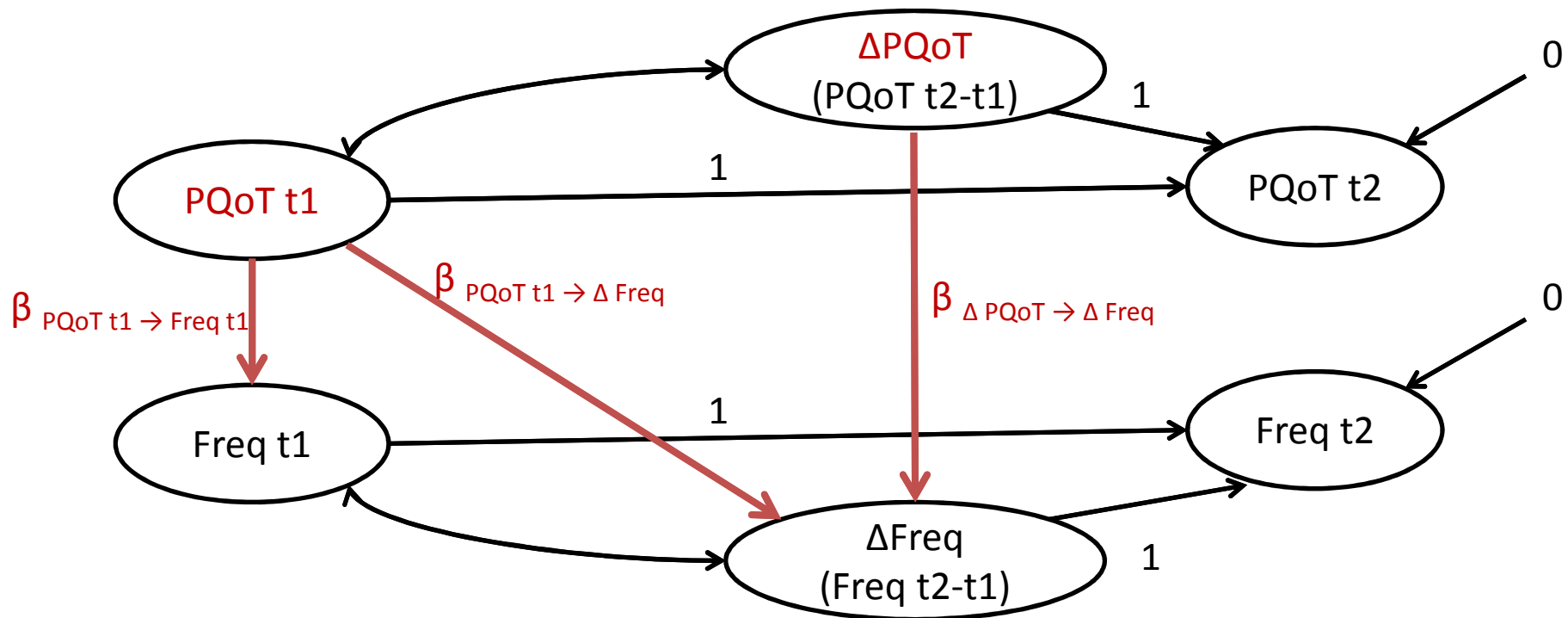
II. … and does the perceived quality of university training support this development?

a. Fit of the models of the impact of PQoT on Freq

	X²	df	RMSEA	NNFI	CFI	sRMR
CD1 „Teaching"	147.5	51	.066	.960	.969	.050
CD2 „Education"	103.6	51	.049	.981	.985	.031
CD3 „Assessment"*	109.2	17	.112	.919	.951	.041
CD4 „Innovation"	190.3	51	.080	.938	.952	.050

- CD1 model: acceptable fit

- CD2 model: good fit

- CD3 model: poor fit, obvious problems, hardly suitable for interpretation

- CD4 model: (borderline) acceptable fit

- CD3 has only two indicators; here, a latent methods factor (M-1) for each aspect was introduced instead of assuming *correlated uniqueness*

II. … and does the perceived quality of university training support this development?

b. Latent parameter values (Estimates) and standardized path coefficients

	M (PQoT t1)	M (Δ PQoT)	β PQoT t1 → Freq t1	β PQoT t1 → Δ Freq	β Δ PQoT → Δ Freq
CD1 „Teaching"	3.37	-.01 ns	.23 ***	-.04 ns	.25 ***
CD2 „Education"	3.38	.11 **	.30 ***	.05 ns	.35 ***
CD3 „Assessment"	2.92	.24 ***	.41 ***	.08 ns	.44 ***
CD4 „Innovation"	2.98	.11 *	.38 ***	.00 ns	.40 ***

1.) The perceived quality of university training is overall poor at t1 (theoretical scale mean is 3.5)…

2.) …but, in the further course of the studies, there is a rise in the perceived quality (except in „Teaching")

II. … and does the perceived quality of university training support this development?

b. Latent parameter values (Estimates) and standardized path coefficients

	M (PQoT t1)	M (Δ PQoT)	β PQoT t1 → Freq t1	β PQoT t1 → Δ Freq	β Δ PQoT → Δ Freq
CD1 „Teaching"	3.37	-.01 ns	.23 ***	-.04 ns	.25 ***
CD2 „Education"	3.38	.11 **	.30 ***	.05 ns	.35 ***
CD3 „Assessment"	2.92	.24 ***	.41 ***	.08 ns	.44 ***
CD4 „Innovation"	2.98	.11 *	.38 ***	.00 ns	.40 ***

3.) Across all competence domains, a high perceived quality of training fosters the frequency of practical application of competences…

4.) …and, if the perceived quality of training is rising in the course of the studies, there is also a positive development of the practical application of competences

5.) However, long term effects of the initially perceived quality of training on later competence development were not found.

"Overall" Summary

- **Question I:** In the "Teaching" and "Innovation" domains, there is a positive longitudinal development in teacher students' competences, but not so in the "Education" domain. (The "Assessment" model should not be interpreted due to very poor model fit)

  – Thus, while teacher students are able to enhance their teaching competences in practice, developing educational competences "in vivo" may be more difficult

- **Question II:** In all domains, a high perceived quality of training contributes to a high initial level and a positive development of competence

  – i.e. students appreciate and feel backed by "good" training

Thank you very much for your attention!

Dr. Christoph Schneider (schneider@uni-landau.de)
Dr. Rainer Bodensohn (bodensohn@uni-landau.de)
University of Koblenz-Landau, Campus Landau, Germany

EERA/ECER Conference in Istanbul, September 2013

Reservoir

- Aims of the KOSTA evaluation framework

Multiple Aims
of the KOSTA Evaluation Framework

KOSTA data aim to

- provide the individual teacher student with feedback concerning strongholds or shortcomings in his or her competence profile

- serve to evaluate the outcomes of teacher education systems or institutions

- contribute to answering scientific research questions

KOSTA data are not

- used for any grading purposes
  (and thus do not leave a „secure environment")

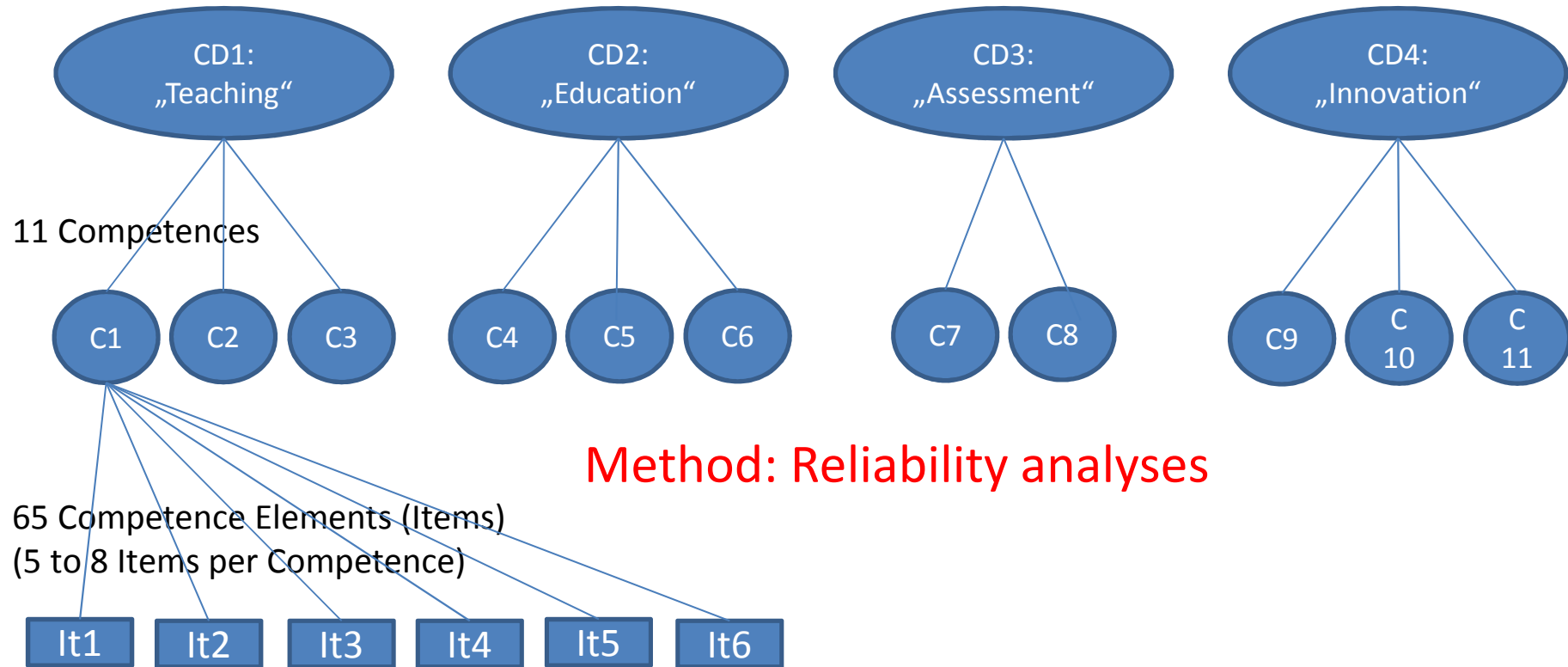   - which, in turn, is supposed to reduce a social desirability bias

Reservoir

- Preliminary Study 1: Scale Reliabilities

Preliminary Study 1

Do the item sets (of competence elements) assigned to each of the eleven competences constitute reliable units of observation?

Preliminary Study 1: Structural Overview

4 Competence Domains

CD1: „Teaching"   CD2: „Education"   CD3: „Assessment"   CD4: „Innovation"

11 Competences

C1 C2 C3   C4 C5 C6   C7 C8   C9 C 10 C 11

Method: Reliability analyses

65 Competence Elements (Items)
(5 to 8 Items per Competence)

It1 It2 It3 It4 It5 It6

… each with ratings of three aspects

Preliminary Study 1:
Do the item sets (of competence elements) assigned to each of the eleven competences constitute reliable units of observation?

- The competence elements were derived from literature; their wording was backed by normative considerations

- Thus, we can not assume that this postulated structure (of exactly eleven competences) might be found „one-to-one" in exploratory structural analyses

- As a minimum requirement, however, what is assigned to *one* competence should form a reliable unit…

- … for each of the three rating aspects

Preliminary Study 1: Do the item sets (of competence elements) constitute reliable units ?

- Cronbach's α of the three aspects for the eleven competences (t2 data)

a) Overall importance
b) Frequency
c) Perceived quality of training
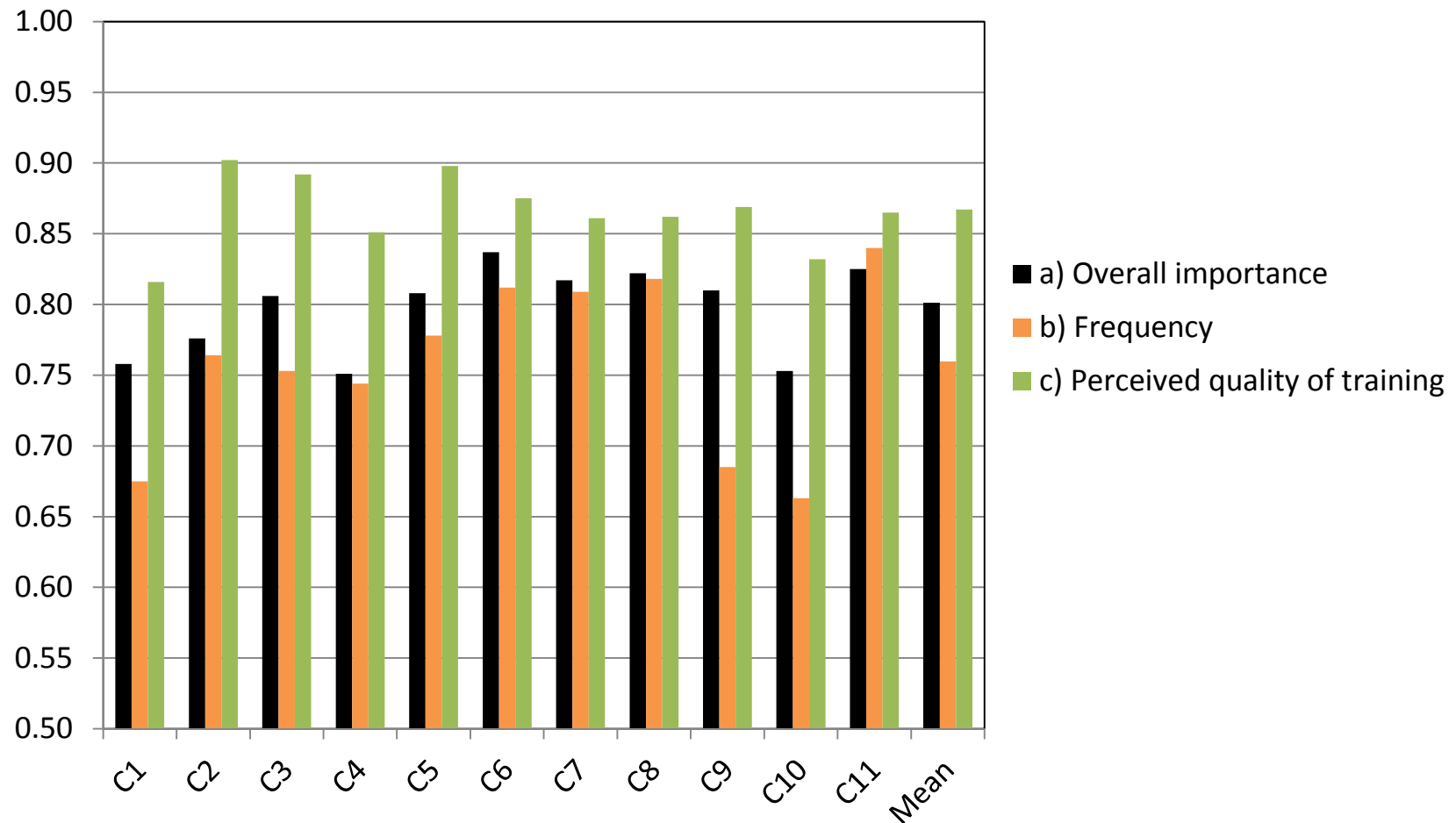
Preliminary Study 1: Do the item sets (of competence elements) constitute reliable units ?

- Consistencies are acceptable to high (nearly) across all competences and aspects

- The self ratings are, to a small degree, more consistent at t2 than at t1 (not shown)

- C9 und C10 from the „Innovation" domain are relatively inconsistent in the frequency aspect

- Summary for preliminary study 1:
On a whole, the observed competence elements seem to form reasonably reliable constructs in the sense of the normatively proposed structure
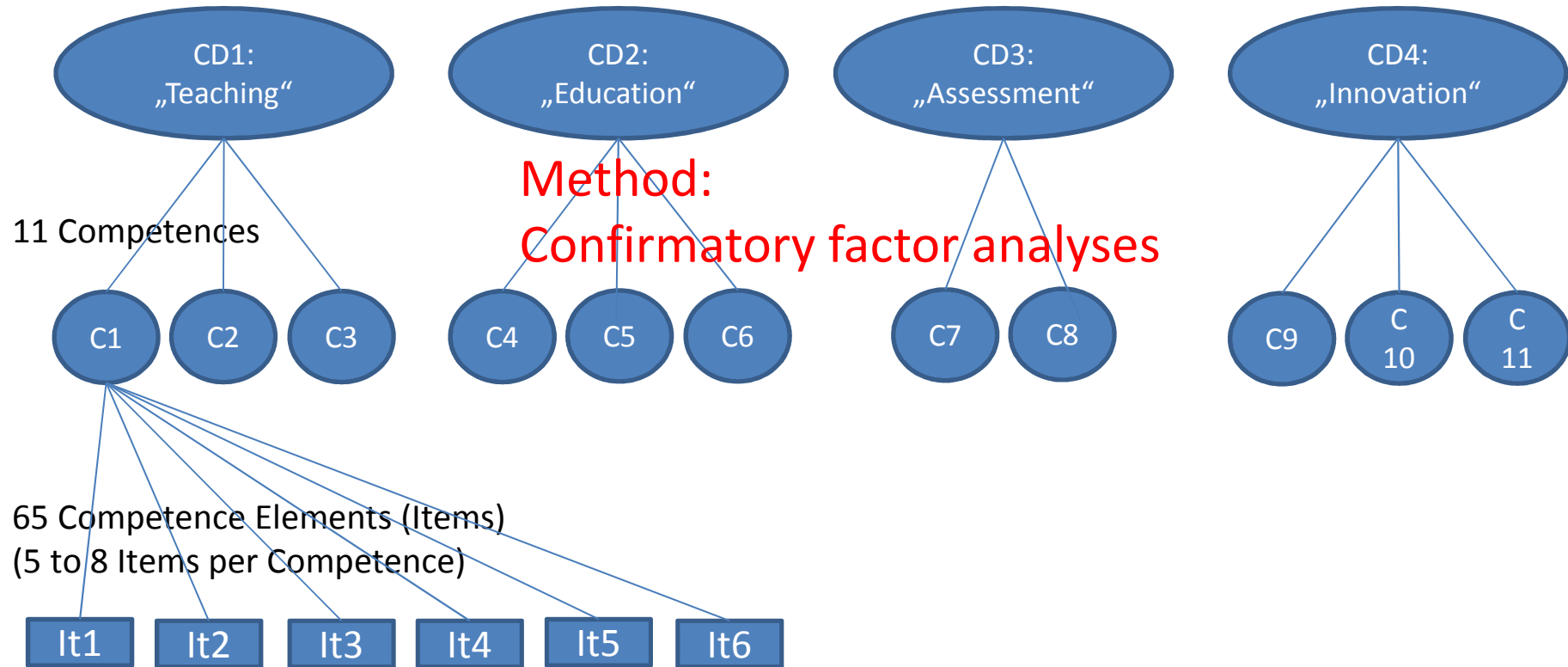
Reservoir

- Preliminary Study 2: Empirical Structure of Competence Domains

Preliminary Study 2

Is it possible to empirically distinguish the four broad competence domains postulated by the KMK?

Preliminary Study 2: Structural Overview

4 Competence Domains

Method:
Confirmatory factor analyses

11 Competences

65 Competence Elements (Items)
(5 to 8 Items per Competence)

... each with ratings of three aspects

Preliminary Study 2: Is it possible to empirically distinguish the four broad competence domains postulated by the KMK?

- For each of the three aspects, the eleven competences were submitted to confirmatory factor analysis

  - M1: general factor model
    -> all eleven competences are indicative of one factor, to be labelled as „general teacher (meta-)competence"

  - M2: orthogonal factors model
    -> the eleven competences are indicative of four *uncorrelated* factors (competence domains)

  - M3: correlated factors model
    -> the eleven competences are indicative of four *correlated* factors (competence domains)

  Based on self report data collected at t2; N = 410

Preliminary Study 2: Is it possible to empirically distinguish the four broad competence domains postulated by the KMK?

„Importance" aspect	X^2	df	RMSEA	NNFI	CFI	sRMR	AIC
M1 (general factor)	365.0	44	.134	.96	.97	.040	409.0
M2 (4 orth. factors)*	1668.3	44	.30	.80	.84	.55	1712.3
M3 (4 corr. factors)	137.3	38	.080	.98	.99	.026	193.3
„Frequency" aspect	X^2	df	RMSEA	NNFI	CFI	sRMR	AIC
M1 (general factor)	571.4	44	.171	.92	.94	.067	615.4
M2 (4 orth. factors)*	1365.3	44	.27	.79	.83	.46	1409.3
M3 (4 corr. factors)	282.7	38	.125	.97	.97	.048	338.7
„Quality of preparation" aspect	X^2	df	RMSEA	NNFI	CFI	sRMR	AIC
M1 (general factor)	586.9	44	.174	.94	.95	.050	630.9
M2 (4 orth. factors)*	1550.7	44	.29	.82	.86	.55	1594.7
M3 (4 corr. factors)	155.98	38	.087	.98	.99	.027	212.0

* model has identification problems

Preliminary Study 2: Is it possible to empirically distinguish the four broad competence domains postulated by the KMK?

- Across all aspects, models behave in a similar fashion:

  o M1 (general factor model) does not display a convincing model fit

  o M2 (orthogonal factors model) -> absurd!

  o M3 (correlated factors model): Not all fit indices are acceptable; nonetheless M3 fits the data (far) better than M1
  (tests of ΔX^2 were also sign.)

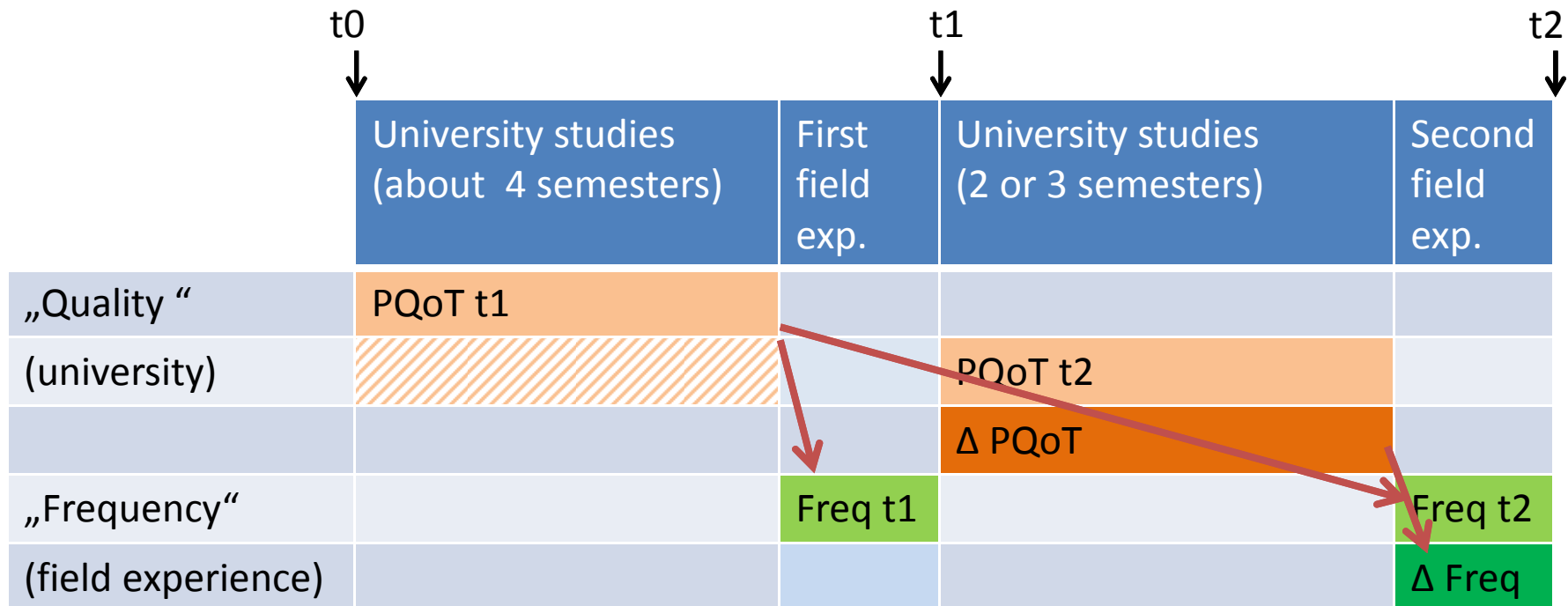- Summary for preliminary study 2:
  The normative assumption that the eleven competences are indicative of four broader competence domains is backed at least partially

- Thus, it seems appropriate to model these four competence domains by the eleven competences as indicators for the purpose of further analysis

Reservoir

- Longitudinal Design Overview

KOSTA: Longitudinal Design Overview

- What do the self ratings in the "Perceived Quality of training" (PQoT) and "Frequency" (Freq) aspects refer to?

	t0		t1		t2
	University studies (about 4 semesters)	First field exp.	University studies (2 or 3 semesters)		Second field exp.
„Quality "	PQoT t1				
(university)			PQoT t2		
			Δ PQoT		
„Frequency"		Freq t1			Freq t2
(field experience)					Δ Freq

- The PQoT measures refer to time intervals *prior* to the field experiences; Freq measures refer to the field experiences *themselves*

- Thus, effects of PQoT in Freq may be interpreted directionally

Reservoir

- The pros and cons of self-ratings

Why apply self ratings in the evaluation of teachers' pedagogical competence (I/IV)?

- The validity of self ratings (a) in general and (b) in assessing teacher behaviour has repeatedly been questioned:

  o Self ratings of teacher behaviour might not be valid evaluations of proper behaviour but may rather express „self concept" (see Hartig & Jude, 2007, p. 25) or „expectations of a rise in competence " (Hascher, 2011, p. 431) .

  o Whenever „hard" criteria or methods of assessing a construct are at hand, self ratings turn out to be relatively poor predictors:

  o In a recent meta analysis, self ratings of „intellect" correlated weakly (.33) with intelligence test results (Freund & Kasten, 2012)

- BUT: Specifically in assessing teacher behaviour, such hard criteria are rarely (if ever) at hand

Why apply self ratings in the evaluation of teachers' pedagogical competence (II/IV)?

- Self ratings of teacher behaviour may not universally possess a high validity, but…

  - … it is questionable whether other sources of information (particularly expert ratings) are generally more valid (Howard, 1980; 1990). Moreover, concordance is poor across perspectives (teacher self ratings, expert ratings, student ratings) is overall poor (Clausen, 2002; Schneider & Bodensohn, 2011).

  - … in contrast to "experts" or external observers in the classroom, only the teacher may subjectively pedagogically substantiate his or her behaviour (see Clausen, 2002)

  - … particularly student ratings referring to teaching quality and teacher behaviour are less complex than "professional" competence models. Thus, outcomes are hardly comparable (Schneider & Bodensohn, 2011)

  - … in evaluating teacher education programmes, self ratings (along with other measures) are a commonly used source of information (Darling-Hammond, 2006; Darling-Hammond et al., 2002).

Why apply self ratings in the evaluation of teachers' pedagogical competence (III/IV)?

- The pros…

  - o If the self rating is collected in a *protected room* (i.e. ratings are not used for grading or rating purposes other than providing individual feedback), social desirability effects are minimized; in the VERBAL (Bodensohn & Schneider, 2008) and REBHOLZ (Schneider & Bodensohn, 2008) projects, all teacher students' self ratings were more negative (i.e. more critical) than ratings by supervising mentors in schools (supposedly the "experts")
    - − Unfortunately, in the Freund and Kasten (2012) meta-analysis, the *secure environment* variable was not included as a potential moderator
  - o Teacher students are often grateful for the individual feedback provided by comparing their self ratings to norm values for their peer group. Students often state that the profile of their self ratings enable them to identify personal shortcomings and subsequently to work on these shortcomings (data source: qualitative statements in counselling sessions with university teacher educators)

Why apply self ratings in the evaluation of teachers' pedagogical competence (IV/IV)?

- The pros…

  o Self ratings are, by far, the most economic approach for assessing competences in large samples or populations. In practical field experience episodes in the course of pre-service university teacher education, individual assessment of the competences of each individual teacher student - for formative evaluation purposes - by means of expert or observer ratings is hardly viable for economic reasons

  o Possibly, the "manner of speaking" in item wording may help to reduce potential response bias. In the KOSTA project (and its predecessors), we decidedly ask for the *frequency* by which a certain behaviour element was applied, not for the perceived *quality* (or the *success*) of the behaviour. We believe that students feel less pressurized when we ask for the frequency . Our rationale is that in a learning setting it is at first important to "try out" a supposedly "adequate" behaviour element – success may come later and depends on personal experience

Reservoir

- What other aspects have been assessed in KOSTA (but not been mentioned in this presentation)?

Originally, each item was rated under five aspects (not just three)

	Label	Wording	
a	**Overall importance**	Which overall importance do I assign to this competence?	(1) very important … (6) completely unimportant
b	**Frequency of application**	How often did I apply this competence?	(1) very often ... (6) never
c	Difficulty of application	How hard did I find it to apply this competence?	(1) very easy ... (6) very hard
d	Desired significance of this competence in university training	Which emphasis should be put on this competence in university training ?	(1) great emphasis ... (6) negligible
e	**Perceived quality of university training**	How well did the university train me towards this competence?	(1) very well ... (6) not at all

Is there empirical reason in assessing all five aspects?

- In a pilot study (Weresch-Deperrois & Bodensohn, 2010), items of the eleven competences (with five aspects each) were submitted to a joint exploratory factor analysis

- In a three factor solution….

  o … all "Frequency" and "Difficulty" (negatively) ratings load on a common factor (the more "difficult" a behaviour element is thought to be, the less it is applied; it is thus dispensable to explicitly ask for "Difficulty")

  o … all "Overall Importance" and all "Desired Significance" ratings form another common factor (subjects are unable to differentiate between the two aspects)

  o … the "Perceived Quality of University Training" makes up a proper factor

- The findings of the pilot study could be backed by confirmatory analyses (not presented here)

To avoid redundancy, two aspects were excluded from further analyses

	Label	Wording	
a	Overall importance	Which overall importance do I assign to this competence?	(1) very important … (6) completely unimportant
b	Frequency of application	How often did I apply this competence?	(1) very often ... (6) never
c	Difficulty of application	How hard did I find it to apply this competence?	(1) very easy ... (6) very hard
d	Desired significance in university training	Which emphasis should be put on this competence in university training ?	(1) great emphasis ... (6) negligible
e	Perceived quality of university training	How well did the university train me towards this competence?	(1) very well ... (6) not at all

Reservoir

- Which conception of competence has been applied in KOSTA (but has not been mentioned in this presentation)?

The roots: Competence as seen by "Deutscher Bildungsrat" (1965-1975)

- Four decades ago, a German federal institution concerned with education and system issues emphasized on the **"Handlungskompetenz"** concept

- The term has no 1:1 equivalent in English language, it may be circumscribed as

  - action competence (its literal translation)

  - ability of an individual to perform

- Thus, "Handlungskompetenz" is more than a disposition

- The term also comprises the readiness to act out and to develop capabilities in practise

- In consequence, "Handlungskompetenz" is best assessed by observing behavioural outcomes

Weinert´s integrating concept of competence

- Weinert bridged the gap between the psychological / pedagogical concepts and sociological concepts of competence

- He defined competence as **"…a roughly specialised system of abilities, proficiencies, or skills that are necessary to reach a specific goal. This can be applied to individual dispositions or to the distribution of such dispositions within a social group or an institution"**

The concept underlying (our) modelling of competences

- In 2009, Josef Leisen, a physics-educationalist explained the term in a more algebraic way:

competence   = knowledge + ability + action

**   = active handling of knowledge**

- Thus, inferences about a person's competences should ideally be based on observations of his or her behaviour in practise