Kristina

**F.1**

**F.2**

**F.3**

**F.4**

\* It is demanded by researchers as well as politics that teachers engage with evidence

\* we know about different definitions of evidence, when we speak about evidence, we mean.... formal (e.g. gradings) & informal data (e.g. observations into the classroom)

\* ... which results in the two concepts of DBDM & RIEP

\* we adress both of these concepts with our studies, the first study is in the context of DBDM and the second study in the context on RIEP

\* both concepts can balance each other and Brown and colleagues combined them...

\* As an example: data gives us a very individual data basis

\* But what we don't have in the data itself is the view from outside of the school.

\* Data use can leave blind spots

\* Advantages of RIEP -> uncover these blind spots

\* Because of research as an external input

\* Unfortunately there are some barriers making it harder for teachers to engage with evidence

- like a lack of competences - data-literacy as well as research-literacy

- not enough time during everyday practice

- considering it as irrelevant, not useful, far away from their practical needs

\* And there are some barriers concerning the correct interpretation, like cognitive biases

**F.5**

**F.6**

- Example Anchoring: Participants in an experiment were asked to estimate how old Gandhi became (Strack & Mussweiler, 1997)

- before their estimations, one group got an anchor of 9 (estimation: 50) and the other of 140 (estimation: 67) ....

- The Anchoring effect has been studied for half a century - nevertheless are the mechanisms behind the effect not completely clarified

- There are different theories existing how anchoring occurs, but these mechanisms behind are not our focus today - so let me tell you about some results of some previous research

- in the long run, it's absolutely rational using Anchors as a helpful estimate

- but overcoming this influence leads to more precise estimations

- and when we want teachers to engage with evidence, data as well as scientific research, we need them to make precise estimations

- This leads to the quote of Tversky and Kahneman, whom have most prominently described anchoring effects...

- "A better understanding of these heuristics and of the biases to which they lead could improve judgments and decisions in situations of uncertainty" (Tversky und Kahneman, 1974, p. 9)

- -> situations of uncertainty are characteristic of the educational systems, of schools and teaching -> like the Anchoring effect in the context of EIP

**F.7**

**F.8**

**F.9**

- We conducted a randomized controlled trial with two experimental groups...

- Participants were guided through 3 virtual experiments

- The task in these experiments was always to sample in a sequential way until the participant was confident that she or he could adequately answer a question on an educational topic by using data

- e.g. "Do girls get better grades than boys"

- Before we go through the whole figure, let me show you concretely how these experiments look

**F.10**

- these are screenshots of each of the three experiments, sorry that they are in german but...

- first experiment as an example for exercise

- participants could sample freely with this open text box: they could enter a number in it

- with the click on the button they got the results in a graphic

- on the basis of this graphic they could decide whether to sample more or if it's enough to answer the question

- so they could enter numbers in the text box as long as they wanted to sample more -> sequential sampling -> graphic update with every new sampling

- same procedure for the second virtual experiment, but with these buttons as given sampling options

- third experiment again with an open text box

**F.11**

- actually the same but with other buttons

**F.12**

- and finally the control group, only with open text boxes

- also in the second experiment

**F.13**

- I have shown you these three experiments

- What you can see here additionally is, that exp. 2 & 3 were about educational topics

- And with these buttons in the second virtual experiment we set the anchor

- With these, we hypothesized that the anchors have an influence on the sampling in the third experiment

- More precisely, on the intital sampling (the first number the praticipants wrote in the open text box) and on the maximum cumulated sampling (all numbers added together)

**F.14**

- with the anchor options as independent variables and the sampling as dependent variables

**F.15**

- We compared as first analyses the sampling of the two anchor conditions

- at the left ... at the right ... with the experimental groups as independant variable in the x-axis and the sampling as dependent variable on the y-axis

- We conducted Bayesian Wilcoxon-tests, with large effects for both of our hypotheses

- BF revealed strong evidence, also for both hypotheses

- -> large anchoring effects on engaging with data

- H1: Our hypothesis is more than a hundred times more likely than the null hypothesis

- wilcoxon-tests and no t-tests because the data is not normally distributed

- H1: exact BF: 848; W = 215; p-value <.001

- H2: exact BF: 40.307; W = 289; p-value <.001

**Kirstin**

**F.16**

- in our second study we again analyzed anchoring effects but this time we analyzed whether anchors influence student teachers' interpretation of scientific evidence

**F.17**

- To answer this research question, we conducted an online experiment with 233 student teachers from the KA University of Eduaction.

- here you can get a quick impression of the demographics of our sample- most important information is that our participants were mostly female and studied at least one STEM subject

**F.18**

- Our study materials included scientific study reports on two different topics from educational research in which either a small, a large or no anchor was included.

- Here you can see an example of such a study report on discovery learning

- Let me explain the material and design in more detail

- In general, the study reports were structured identically:

- They first informed about

- the research question - a group of educational researchers was interested in,

- then about the experimental design of the study with one experimental and one control group

- as well as the drawn sample size to investigate the research question.

- Each student teacher randomly received two different study reports where the drawn sample size also varied randomly - either a small sample size of 15 participants or a large sample size of 500 participants per group.

- This approach resulted in

- two control groups, where student teachers were presented with the same sample size in both study reports, so no anchor was set, e.g., two times 15 participants

- and two experimental groups either with a small or a large anchor. For example the first EG was presented with a small anchor - a sample size of 15 and then a large sample size of 500 participants

**F.19**

- After reading the study report, participants both in the ExperimentalGs and CGs had to answer different questions.

- For today's presentation only one of these questions is important, namely the one right at the bottom outlined in green.

- Referring to this, the student teachers had to decide whether ‘the number of sampled students is appropriate to answer the research question'

- they had to give their answer on a Likert-scale ranging from totally disagree to totally agree

**F.20**

- Using these materials, we in general assumed that the sample size in study 1 acts as an anchor and therefore influences the appropriateness ratings of the sample size in study 2.

- Before I explain our hypothesis in more detail, just a quick reminder which group is which because this can be a bit confusing but is important to understand our hypothesis:

Experimental group 1 is the group with a small anchor - meaning participants were first presented with a sample size of 15 participants and then with 500 participants per group. In contrast, the experimental group 2 is the group with the large anchor first 500, then 15 participants.

- More specifically, we formulated a hypothesis that consisted of two parts.

- We assumed that the experimental group 1 - the group with the small anchor - rates the presented sample size 2 as more appropriate than the experimental group 2 with the large anchor

- and on the other hand that there is no difference in the appropriateness rating of sample size 2 between our control groups.

- To test this hypothesis we conducted a Bayesian ANOVA.

- And this led to the following results:

**F.21**

- on the y-axis you can see the sample size in Study 1, right at the top you can see whether the sample size in study 2 consisted of 15 or 500 participants per group and on the x-axis you can see the adjusted appropriateness of sample size 2

- Hence, you can see the results of the ExperimentalG1-the group with the small anchor- in the bottom right corner, the results of the EG2-the group with the large anchor- in the top left corner.

- Just a quick side note: To create the dependent variable adjusted appropriateness rating of sample size 2, we subtracted the appropriateness rating of the first study sample from the appropriateness rating of the second one.

- It is quite obvious that in line with our hypothesis, the experimental groups differ in their appropriateness ratings: EG 1 - in the bottom right corner - rates the sample size 2 as more appropriate than EG 2 - top left corner. It can be clearly seen that this difference indicates a large effect (d-EGs: 2.97)

- In contrast,but again in line with our hypothesis, the control groups in the left corner at the bottom and right corner at the top barely differ in their appropriateness ratings of sample size 1 and sample size 2 (d-CGs: 0.29).

- Bayes Factors also indicate strong evidence for our hypothesis

**F.22**

I will conclude our presentation by discussing the results of both studies.

**F.23**

- In both studies, we found evidence for anchoring effects in the context of EvidenceISchoolP. While Study 1 shows evidence for anchoring effects on teacher students’ engagement with data, the second one presents evidence for anchoring effects on interpreting scientific evidence.

- But when we think of the practical realization of EISP in general, what consequences can occur because of the anchoring effect?

**F.24**

- When we think of school practice there are many situations were anchoring can occur, for example in grading situations or in statewide assessments were benchmarks can act as an anchor and lead to different practical decisions depending on which anchor has acted. But let me illustrate another example in more detail

- In the context of EIP, large samples of large-scale assessments such as PISA could act as an anchor in the evaluation of experimental studies. The large PISA-samples could lead to a devaluation of the quality and validity of experimental studies due to their comparatively small samples. In turn, teachers might take less account of findings form such experimental studies in their school practice. But we all know that sample size is not the only indicator of quality and validity.

- Benchmarks that are integrated in such

- This example illustrates that although EIP is supposed to be a more objective practice, it is not free of distortions / biases such as anchoring effect. There might also be further biases such as confirmation bias - Kim Schildkamp already talked about in her keynote.

- It is important to better understand these biases in the context of EvidenceISP to improve teachers' engagement or more precisely interpretation of evidence.

- Our next steps are to find approaches to take these results into account in order to foster teachers' engagement with evidence.

- And with this statement I will close our presentation – Thank you for your attention

We are looking forward to your questions and the discussion