Universität Zürich[UZH]

Algorithm Aversion by Teachers?

Esther Kaufmann
University of Zurich, Switzerland

David V. Budescu
Fordham University, NY, US

SJDM Conference 2019, November 16

Kaufmann, E., & Budescu, D. V. (2019). Do Teachers Consider Advice? On the Acceptance of Computerized Expert Models. *Journal of Educational Measurement.*

FNSNF SWISS NATIONAL SCIENCE FOUNDATION
The National Science Foundation (SNSF) funded a research visit for collaborative research with Professor D. V. Budescu, Fordham University (US) to conduct this research project.

Universität Zürich[UZH]

Teachers' Judgment and Decision-Making

- Daily judgments, e.g., on students' performance and motivation
- Different meta-analyses: Hoge and Coladarci (1989); Kaufmann, Reips, & Wittmann (2013); Südkamp, Kaiser, and Möller (2012) summarized in Kaufmann 2019 SJDM Poster.
  - Moderate judgment achievements with room for improvement.

➡ Are (computerized) expert models (e.g., algorithm, bootstrapping models) helpful in improving teachers' judgment and decision-making?

Universität Zürich[UZH]

Recent Reviews

Recent reviews covering the topic of the success of expert models

Meta-analysis	Inclusion criteria
Grove et al. (2000)	Human outcome – medical and psychological tasks
Aegisdottir et al. (2006)	Human outcome – counselling tasks
Armstrong (2001)	No criterion restrictions
Lens-Model based	
Camerer (1981)	No criterion restrictions
Karelaia and Hogarth (2008)	No criterion restrictions
Kuncel, Klieger, Connelly, & Ones (2013)	Academic and work performance settings
Kaufmann, Reips & Wittmann (2013)	No criterion restrictions

See Kaufmann, E., & Wittmann W. W. (2016). The success of linear bootstrapping models: Decision domain-, expertise-, and criterion-specific meta-analysis. *PLoS ONE 11(6):* e0157914.

Universität Zürich[UZH]

Recent Reviews

Recent reviews covering the topic of the success of expert models

Meta-analysis	Inclusion criteria
Grove et al. (2000)	Human outcome –
Aegisdottir	
Armstrong (	
Lens-Model	
Camerer (19	
Karelaia an	
Kuncel, Klieger, Connelly, & Ones (2013)	Academic and work performance settings
Kaufmann, Reips & Wittmann (2013)	No criterion restrictions

Psychometric meta-analysis (Kaufmann & Wittmann, 2016):

The success of expert models is underestimated

due to missing study artifact corrections (e.g., measurement error).

See Kaufmann, E., & Wittmann W. W. (2016). The success of linear bootstrapping models: Decision domain-, expertise-, and criterion-specific meta-analysis. *PLoS ONE 11(6):* e0157914.

Universität Zürich[UZH]

Recent Reviews

Recent reviews covering the topic of the success of expert models

Meta-analysis	Inclusion criteria
Grove et al. (2000)	Human outcome – medical and psychological tasks
Aegisdottir et al. (200	asks
Armstrong (2001)	
Lens-Model based	
Camerer (1981)	No criterion restrictions
Karelaia and Hogarth (2008)	No criterion restrictions
Kuncel, Klieger, Connelly, & Ones (2013)	Academic and work performance settings
Kaufmann, Reips & Wittmann (2013)	No criterion restrictions

Open question:

Do people really accept the advice of expert models?

See Kaufmann, E., & Wittmann W. W. (2016). The success of linear bootstrapping models: Decision domain-, expertise-, and criterion-specific meta-analysis. *PLoS ONE 11(6):* e0157914.

Acceptance of Algorithm Advice?

- Burton, Stein and Jensen (2019): 61 Studies (1950-2018)
- Studies vary widely:
  - Participants are often students, seldom real experts or a comparison of expert vs. non-experts
  - Different fields

➡ The acceptance of expert models varies widely (see e.g., Dietvorst et al., 2014; Dietvorst & Bharti, 2019; Longoni et al., 2019; Önkal et al., 2009; Yeomans et al., 2019 vs. Germann & Merkle, 2019; Logg et al., 2019)

➡ Do teachers really accept the advice of expert models?

Universität Zürich[UZH]

Open Questions

- Does providing teachers with advice from **an expert model and/or a school** counselor affect their decisions? (Based on their behavioral judgments and stated preferences in a survey)

- Does the teacher's acceptance of advice depend on **task attributes** (e.g., number of advice sources, task difficulty)?

➡ Experiments (advice given vs. advice asked for)

Judgment Task

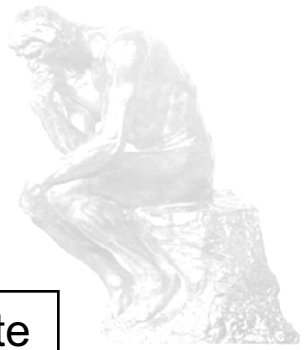
Student A: Mary	Characteristics	Student B: Tom
female	**sex**	male
none	**problems**	none
member of the tennis team	**extracurricular activities**	member of an art and drama group
received no support	**previous student needs**	received no support
	grades	
F	**math**	F
D	**science**	D
C	**english**	C
D	**social studies**	D

Hard tasks (2 out of 6)
Students who had **similar math grades** (“balanced”)

Easy tasks (4 out of 6):
Students who **varied in their math grades** (“unbalanced”)

Universität Zürich[UZH]

Judgment Task(s): Receiving Advice



DV's:
Judgment on students' potential

Teachers have to judge which student they would promote with additional math tutoring hours. **A computerized expert model recommends promoting student A.**

Manipulation

Advice source
Number of advice sources
Advice type

Judgment on certainty

Teachers are also asked how certain they are in their judgments.

Sample: 164 volunteer Swiss middle and high school teachers.

Results: Teachers follow Advice

Advice source	N	Task difficulty: Easy ($K = 4$)	Task difficulty: Hard ($K = 2$)	Overall	
Two agreeing	35	0.60 (0.19)	0.74 (0.35)	0.65 (0.18)	
School counselor	38	0.61 (0.19)	0.64 (0.36)	0.62 (0.20)	***
Expert model	31	0.56 (0.14)	0.56 (0.35)	0.56 (0.16)	***
Two disagreeing	33	0.48 (0.10)	0.41 (0.36)	0.46 (0.15)	
Total	137	0.56 (0.17)	0.59 (0.37)	0.58 (0.19)	

Note. In the case of two conflicting sources the values represent the number of cases where the teachers followed the advice of the expert model.

- A two-way ANOVA with one between-subjects factor (Advice source) and one within-subject factor (Task difficulty), shows **a significant effect of source** ($F(3, 137) = 7.35$, $p < .00$).
- Tukey's post hoc tests reveal that, in every case, teachers follow significantly **more advice from the school counselor and the two agreeing sources than the two disagreeing sources**. There was no effect for task difficulty ($F(1, 133) = 0.55$, $p = .46$)

Results: Teachers' Judgment on Certainty

Advice source	N	Task difficulty: Easy (K = 4)	Task difficulty: Hard (K = 2)	Overall	
Two agreeing	35	5.85 (1.82)	5.37 (1.99)	5.69 (1.80)	
Expert model	31	5.97 (1.80)	4.50 (2.10)	5.48 (1.78)	
Two disagreeing	33	5.83 (1.58)	4.80 (1.49)	5.48 (1.39)	***
School counselor	38	5.54 (1.57)	4.05 (2.12)	5.04 (1.57)	
Control	26	5.26 (2.24)	3.34 (2.48)	4.62 (2.11)	
Total	163	5.70 (1.79)	4.46 (2.12)	5.29 (1.74)	

Note. Judgment certainty could range from 1 (*very uncertain*) to 9 (*very certain*).

- A two-way ANOVA with one between-subjects factor (Advice source) and one within-subject factor (Task difficulty), shows a **significant effect for source** ($F(4, 158) = 2.51$, $p < .05$).
- Post-hoc tests show that teachers feel **significantly more confident** if they got advice from two **agreeing advice sources than in the control condition** (without any advice).
- **No interaction** between **advice source and task difficulty** ($F(4, 158) = 3.74$, $p = .09$).

Summary: Experiment 1

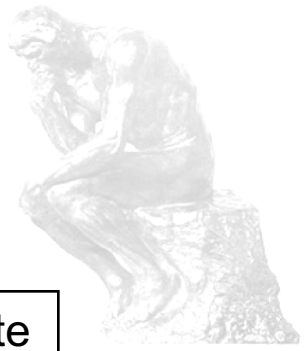
- Teachers selected the recommended student more often than expected by chance (**considering advice**) -> especially the advice of human counselors.
- Surprisingly, their **certainty is affected more by the expert models** than the human forecasters, although this difference is not significant.

Do our results misjudge teachers' receptiveness for advice?
Daily school business teachers have to seek adivce actively….

Experiment 2

Universität Zürich[UZH]

Judgment Task(s): Asking for Advice



DV's:
Judgment on students' potential

Teachers have to judge which student they would promote with additional math tutoring hours.

- School counselor
- Computerized expert model
- Both, computerized expert model and a school counselor

A computerized expert model recommends promoting student A.

Manipulation

Advice source
Number of advice sources

Summary: Experiment 2

- Sample:
  - 99 Inservice teachers (Middle and high school teachers)
  - 63 Preservice teachers (Middle school teacher *students*)
- **Results:**
  - Teachers use **advice selectively** (72.3% of judgment tasks without advice)
  - Teachers ask for **advice in hard** significantly more than in easy tasks (*k*) (Inservice: $\chi^2(1, K = 396) = 64.6$, $p = .00$, Preservice: $\chi^2(1, K = 252) = 67.4$, $p = .00$).

Preference of Advice Source

		Favorite source of single advice (%, *k*)	
Teachers	*k*	School Counselor	Expert Model
Inservice	78	83.3 (65)	16.7 (13)
Preservice	59	81.4 (48)	18.6 (11)
Total	137	82.5 (113)	17.5 (24)

Note. *k* = the number of asked advice across tasks

They chose the **advice of a school counselor** significantly more frequently than that of an expert model (Inservice: $\chi^2(1, k = 78) = 34.6$, $p = .00$; Preservice: $\chi^2(1, k = 59) = 23.2$, $p = .00$).

Preference of Advice Source

Overall Teachers	k	Tasks	Favorite source of single advice (%, k) School Counselor	Expert Model
	22	Easy	72.7 (16)	27.3 (6)
	115	Hard	84.3 (97)	15.7 (18)
Total	137		82.5 (113)	17.5 (24)

Note. k = the number of asked advice across tasks

Preference of Advice Source

Teachers	k	Tasks	Favorite source of single advice (%, k) School Counselor	Expert Model
Inservice		Easy	76.9 (10)	23.1(3)
		Hard	84.6 (55)	15.4 (10)
	78		83.3 (65)	16.7 (13)
Preservice		Easy	66.7 (6)	33.3 (3)
		Hard	84.0 (42)	16.0 (8)
	59		81.4 (48)	18.6 (11)
Overall Teachers		Easy	72.7 (16)	27.3 (6)
		Hard	84.3 (97)	15.7 (18)
Total	137		82.5 (113)	17.5 (24)

Note. k = the number of asked advice across tasks

Especially in hard tasks the advice of a school counselor is preferred (Caution: Small sample size).

Ratings

FORDHAM UNIVERSITY

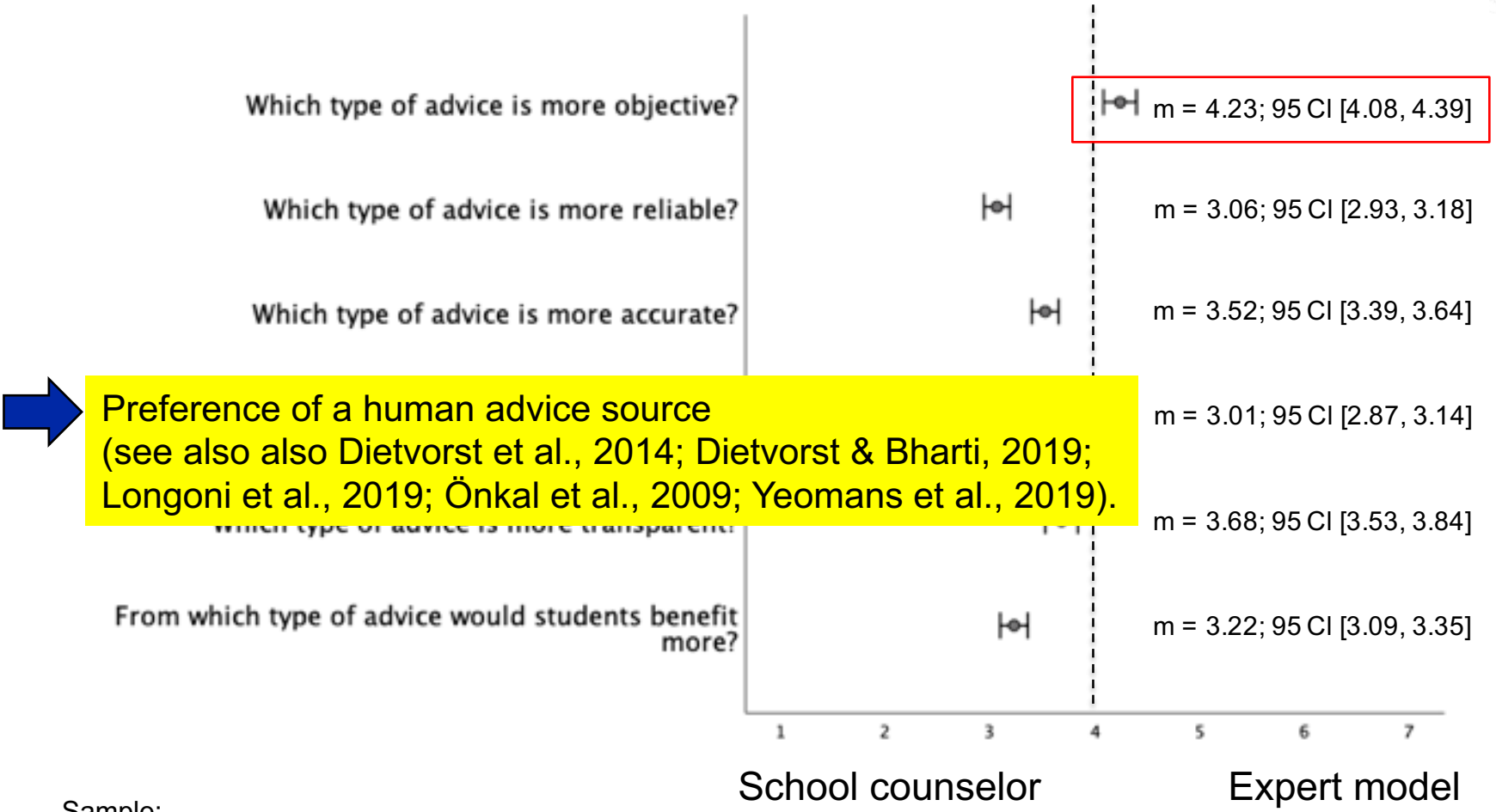
Universität Zürich[UZH]

When comparing the advice of a computerized expert system to that of a school counselor (dean), which of the two would be more **objective** in your opinion?

3 - School Counselor (Dean)	2	1	Both Equally Objective	1	2	Computerized Expert System - 3

>>

Ratings



Preference of a human advice source
(see also also Dietvorst et al., 2014; Dietvorst & Bharti, 2019; Longoni et al., 2019; Önkal et al., 2009; Yeomans et al., 2019).

Sample:
87% Inservice teachers (20% US, 80% Swiss)
13% Preservice teachers (100% Swiss)
Total: 498 (*N*)

Overall Summary

- Teachers followed advice if they got it (when two advice sources agreed)
- Task difficulty: In- and preservice teachers mostly asked for advice in hard tasks.
- Overall: In- and preservice teachers favor advice from human sources over computerized expert models (see also Dietvorst et al., 2014; Dietvorst & Bharti, 2019; Longoni et al., 2019; Önkal et al., 2009; Yeomans et al., 2019).
- Underlying mechanisms?

Limitations and Future Research

- Main restrictions are in the sample (only Swiss middle and high school teachers), tasks (potential on math tutoring), domain (education), online experiment (external validity is critical).

Outlook

- Further studies: Algorithm knowledge by teachers?
- Intervention studies to increase the acceptance of advice, especially of expert models (see Dietvorst et al., 2016).

Universität Zürich[UZH]

Thank you

Esther Kaufmann (University of Zurich, CH)
David V. Budescu (Fordham University, US)

Questions:

esther.kaufmann@ife.uzh.ch /
budescu@fordham.edu

Kaufmann, E., & Budescu, D. V. (2019). Do Teachers Consider Advice? On the Acceptance of Computerized Expert Models. *Journal of Educational Measurement*.

Additional literature on ResearchGate

FNSNF
Swiss National Science Foundation
The National Science Foundation (SNSF) funded a research visit for collaborative research with Professor D. V. Budescu, Fordham University (US) to conduct this research project.