

## **Automatisierung in der Eignungsdiagnostik: Wie Algorithmen die Eignungsdiagnostik verändern**

Die fortschreitende Automatisierung verändert eignungsdiagnostische Prozesse. So können beispielsweise Personalauswahlentscheidungen und Leistungsbewertungen auf Basis automatisch ausgewerteter Daten getroffen werden. Automatisierten Systemen, die für solche Aufgaben genutzt werden, unterliegen häufig moderne Methoden des Machine Learning. Die daraus entstehenden Systeme werden dann oftmals unter dem Begriff "Künstliche Intelligenz (KI)" in der Öffentlichkeit kontrovers diskutiert.

Statt KI für Managementaufgaben jedoch als gänzlich neuartiges Thema anzusehen, kann Forschung aus der Arbeits- und Organisationspsychologie sowie aus dem Bereich Human Factors Eindrücke darüber vermitteln, wie sich Human Resource Management (HRM)-Prozesse durch Automatisierung verändern könnten. Automatisierung betrifft nach Parasuraman et al. (2000) vier generelle Funktionen, die vormals alleine von Menschen übernommen wurden: Informationsakquise, Informationsanalyse, Entscheidungsauswahl und Handlungsausführung. Beispielhaft können diese Funktionen anhand der Personalauswahl mittels Bewerbungsunterlagen klargemacht werden. Nutzt man Automatisierung zur Informationsakquise kann ein automatisiertes System beispielsweise helfen, Informationen aus einer Großzahl an Bewerbungsunterlagen zu sammeln. Informationsanalyse stellt eine weitere Funktion der Automatisierung dar. Beispielsweise können spezifische Informationen aus Bewerbungsunterlagen automatisiert extrahiert und für jeden Bewerbenden zusammengefasst werden. Geht es um Entscheidungsauswahl, kommt gegebenenfalls das vielzitierte Machine Learning zur Anwendung. Hierbei gibt es drei Basisschritte für die Entwicklung von Machine Learning-Algorithmen: Training – Validierung – Testung. Im Fall von Bewerbungsunterlagen könnten Machine Learning-Algorithmen zunächst mit Daten aus vorangehenden Bewerbungsprozessen „antrainiert“ werden, mit deren Hilfe das System Zusammenhänge mit Erfolgskriterien (z.B. Erfolg in Bewerbungsgesprächen, Berufserfolg bei Einstellung) lernen soll. Um sogenanntes „Overfitting“ zu vermeiden (eine zu große Anpassung der antrainierten Algorithmen auf die Spezifika der historischen Daten, was die Generalisierbarkeit der Algorithmen untergräbt), muss darauf geachtet werden, dass trainierte Algorithmen mit bisher „ungesehenen“ Daten kreuzvalidiert werden. Nach der Auswahl eines Algorithmus mit Hilfe der Validierungsstichproben, sollten Algorithmen kontinuierlich anhand neuer Daten getestet werden, um zu evaluieren, ob Vorhersagen (im hier präsentierten Beispiel die Vorauswahl der Bewerbenden) valide sind. Die Validität der Vorhersagen ist hierbei natürlich nur ein Kriterium, zeitgleich sollte beispielsweise auch auf Fairness in den Vorhersagen geachtet werden, da sich bei der Nutzung historischer Daten schnell Qualitätsprobleme der Datengrundlage (z.B. unfaire ungleiche Geschlechtsverteilungen) in den Vorhersagen der antrainierten Algorithmen abbilden können. Unter automatisierter Handlungsausführung als letzte Funktion der Automatisierung, könnte man sich im vorliegenden Beispiel vorstellen, dass die vom System vorausgewählten Bewerbenden direkt durch das System für eine nächste Bewerbungsrunde eingeladen werden.

Die genannten Funktionen der Automatisierung sind zwar nicht nur für die Eignungsdiagnostik attraktiv, tatsächlich scheinen aber vor allem Rekrutierung und Personalauswahl die Gebiete zu sein, in denen Automatisierungsprozesse im Management basierend auf einer Kombination von verschiedenen technologischen Neuerungen

## 10. Assessment Kongress

flächendeckend eingesetzt werden. Eine Vielzahl von Start-ups, aber auch multinationale Konzerne haben sich der Automatisierung dieser Aufgaben verschrieben. Automatisierte Suche und Ansprache von geeigneten Bewerbenden, automatisierte Schaltung von Bewerbungsanzeigen basierend auf Social Media-Profilen und Internetnutzungsverhalten von potentiellen Bewerbenden sowie automatisiert durchgeführte und ausgewertete Bewerbungsgespräche sind nur einige der Beispielanwendungen (teil-)automatisierter Systeme in Rekrutierung und Personalauswahl.

Bei aller Unsicherheit über die zukünftige Entwicklung von automatisierten Systemen und "künstlicher Intelligenz" ist Eines klar: Die Eignungsdiagnostik wird sich durch die fortschreitende Automatisierung verändern. Statt jeden Schritt im eignungsdiagnostischen Prozess selbst durchzuführen, müssen Entscheidungstragende im Falle von automatisierter Unterstützung entscheiden, ob sie automatisch vorgefilterten und bewerteten Daten vertrauen, oder ob sie diese Daten hinterfragen wollen. Im Vertrauensfall erspart die automatisierte Bewertung Arbeit und ermöglicht Freiraum für die Erfüllung anderer Aufgaben. Gleichzeitig führt das aber eventuell zu Unsicherheit über die Güte der endgültigen Entscheidung. Zudem entstehen neue Probleme, wenn man sich auf die Empfehlungen eines automatisierten Diagnostiksystems verlässt: Welche Begründung für eine Ablehnung würde man Bewerbenden geben, die auf Basis der Empfehlungen des Systems aussortiert wurden? Welche Begründung Mitarbeitenden, denen Karriereoptionen verwehrt werden durch automatisierte Leistungsbewertungen? Lehnen Entscheidungstragende automatisierte Empfehlung ab, müssen sie voraussichtlich gut begründen können, warum – wieso sollte man sonst ein kostspieliges, automatisiertes System einsetzen, wenn es Entscheidungstragende dann nicht nutzen? Auch auf der Seite derjenigen, die bewertet werden, verändert die Einführung von automatisierten Entscheidungsunterstützungssystemen die Eignungsdiagnostik. Mussten beispielsweise Bewerbende bisher andere Menschen im Bewerbungsgespräch beeindrucken, müssen sie heute zusätzlich ein automatisiertes Personalauswahlsystem von sich überzeugen.

Es scheint naheliegend, dass sich die fortschreitende Automatisierung von eignungsdiagnostischen Prozessen auf Entscheidungstragende und Entscheidungsempfangende auswirkt. Klassische Forschungstraditionen wie zum Beispiel Forschung zur Nutzung mechanischer Entscheidungsfindung bieten potentiell übertragbare Erkenntnisse. Burton et al. (2020) fassten beispielsweise die Forschung zu mechanischer zusammen und zeigten, dass Menschen ungern Empfehlungen von Systemen nutzen, selbst wenn Systeme im Zweifel bessere Vorhersagen treffen als die Person selbst. Jedoch gibt es auch Erkenntnisse, die genau das Gegenteil gefunden haben – dass Menschen sich eher auf Systemempfehlungen verlassen als auf Empfehlungen von anderen Menschen (Logg et al., 2019).

Forschung, die sich explizit mit automatisierter Eignungsdiagnostik beschäftigt, untersucht vor allem drei Themen: Akzeptanz, Validität und Fairness. Im Bereich **Akzeptanz** deutet die Forschung darauf hin, dass Bewerbende negativ auf automatisierte Personalauswahl reagieren (Acikgoz et al., 2020; Gonzalez et al., 2019; Langer et al., 2019, 2020; Lee, 2018). Allerdings gibt es auch Forschung, die nahelegt, dass auch positive Reaktionen zu erwarten sind, was zeigt, dass Unterschiede zwischen Bewerbenden (z.B. bisherige Diskriminierungserfahrung) und das Design des jeweiligen Diagnostiksystems einen Einfluss auf die Akzeptanz automatisierter Personalauswahl haben (Marcinkowski et al.,

## 10. Assessment Kongress

2020). Besser akzeptiert werden tendenziell Verfahren, die früher im Prozess eingesetzt werden (z.B. zur automatisierten Bewertung von Bewerbungsunterlagen) und solche, in denen nicht vollautomatisiert entschieden wird, d.h. bei denen Menschen das letzte Wort in der Entscheidung haben (für einen Überblick zur Forschung bezüglich Reaktionen auf automatisierte Entscheidungen siehe Langer & Landers, 2021). Des Weiteren deuten erste Ergebnisse darauf hin, dass alleine die Vorstellung einer automatisierten Bewertung von Interviews das Verhalten von Bewerbenden beeinflussen kann (Langer et al., 2020). Erste Forschung untersucht die **Validität** automatisierter Bewertungen von Motivationsschreiben (Campion et al., 2016) oder Bewerbungsgesprächen (Hickman et al., 2021). Allerdings ist hier noch viel Forschung nötig, um Aussagen über die Validität automatisierter Verfahren treffen zu können. In anderer Forschung wird die **Fairness** automatisierter Verfahren thematisiert. Hier wird das Problem **unfairer Biases** in der automatisierten Bewertung betont und gleichzeitig darauf hingewiesen, dass Anbieter von automatisierten Lösungen häufig als Verkaufsargument nutzen, dass ihre Lösungen fairer sein sollen als menschliche Bewertungen (Raghavan et al., 2020). In einer eigenen Forschungsarbeit zeigen wir mögliche Folgen auf: Menschen reagieren weniger stark auf Unfairbehandlung durch automatisierte Systeme, was dazu führen könnte, dass diese weniger wahrscheinlich erkannt wird (Langer et al., 2022).

Zusammengefasst zeigt die bisherige Forschung Effekte der Automatisierung in der Eignungsdiagnostik auf Entscheidungstragende und Entscheidungsempfangende. Aktuelle Trends deuten darauf hin, dass die weitere Entwicklung der Automatisierung von Managementaufgaben in einem Spannungsfeld fortschreiten wird. Auf der einen Seite drängen weitere Neuerungen auf den Markt, die verschiedene Aufgaben zu bewältigen versprechen und die Bundesregierung bläst zu einer Offensive zur Künstlichen Intelligenz. Auf der anderen Seite wird „ethische“, „menschzentrierte“ und „vertrauenswürdige“ künstliche Intelligenz gefordert (<https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>). Auch Implikationen der Datenschutzgrundverordnung und weiterer entstehender Rechtsvorschriften (z.B. der AI Act der Europäischen Kommission <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>) müssen bei der Automatisierung von eignungsdiagnostischen Prozessen bedacht werden. Abschließend bleibt zu sagen, dass die Evolution der Automatisierung der Eignungsdiagnostik ein spannendes Feld ist, auf das Mitglieder\*Innen des Forum Assessments aktiv einwirken können.

### Zentrale Referenzen

- Burton, J. W., Stein, M., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, 33(2), 220–239. <https://doi.org/10.1002/bdm.2155>
- Campion, M. C., Campion, M. A., Campion, E. D., & Reider, M. H. (2016). Initial investigation into computer scoring of candidate essays for personnel selection. *Journal of Applied Psychology*, 101(7), 958–975. <https://doi.org/10.1037/apl0000108>
- Hickman, L., Bosch, N., Ng, V., Saef, R., Tay, L., & Woo, S. E. (2021). Automated video interview personality assessments: Reliability, validity, and generalizability investigations. *Journal of Applied Psychology, Advance Online Publication*. <https://doi.org/10.1037/apl0000695>
- Langer, M., König, C. J., Back, C., & Hemsing, V. (2022). Trust in artificial intelligence: Comparing trust processes between human and automated trustees in light of unfair bias. *Journal of Business and Psychology, advance online publication*. <https://doi.org/10.1007/s10869-022-09829-9>
- Langer, M., & Landers, R. N. (2021). The future of artificial intelligence at work: A review on effects of decision automation and augmentation on workers targeted by algorithms and third-party observers. *Computers in Human Behavior*, 123, 106878. <https://doi.org/10.1016/j.chb.2021.106878>
- Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. *Proceedings of the 2020 FAT\* Conference on Fairness, Accountability, and Transparency*. <https://doi.org/10.1145/3351095.3372828>



BEITRAG 5-DIG

**AUTOMATISIERUNG IN DER  
EIGNUNGSDIAGNOSTIK** WIE ALGORITHMEN DIE  
EIGNUNGSDIAGNOSTIK VERÄNDERN  
DR. MARKUS LANGER

**EXKLUSIV** KI zur Persönlichkeitsanalyse

Fragwürdige Personalauswahl mit  
Algorithmen

## Challenges and benefits of hiring algorithms

Wenn Algorithmen bei der Personalauswahl helfen

Algorithmen bei der Personalauswahl

Die neue Macht der Maschinen

**Robo Recruiting – Dank Algorithmen bessere Mitarbeiter:innen  
finden?**

**What Hiring Algorithms Mean for Job  
Searchers Like You**

## ABLAUF

- Grundlagen
- Automatisierung in der Eignungsdiagnostik am Beispiel HireVue
- Forschung zu automatisierter Eignungsdiagnostik
- Ausblick



Dr. Markus Langer

- Wissenschaftlicher Mitarbeiter am Lehrstuhl für Arbeits- und Organisationspsychologie an der Universität des Saarlandes
- Forschungsschwerpunkte:
  - Einsatz automatisierter Systeme für die Entscheidungsfindung (HRM, Medizin)
  - Erklärbare Künstliche Intelligenz
  - Arbeitsgestaltung beim Einsatz automatisierter Systeme
- Co-Gründer von Algoright e.V.



## ABLAUF

- Grundlagen
- Automatisierung in der Eignungsdiagnostik am Beispiel HireVue
- Forschung zu automatisierter Eignungsdiagnostik
- Ausblick

EXKLUSIV KI zur Persönlichkeitsanalyse

### Fragwürdige Personalauswahl mit Algorithmen

## Challenges and benefits of hiring algorithms

Wenn Algorithmen bei der Personalauswahl helfen

Algorithmus erkennt Persönlichkeit,

Algorithmen bei der Personalauswahl

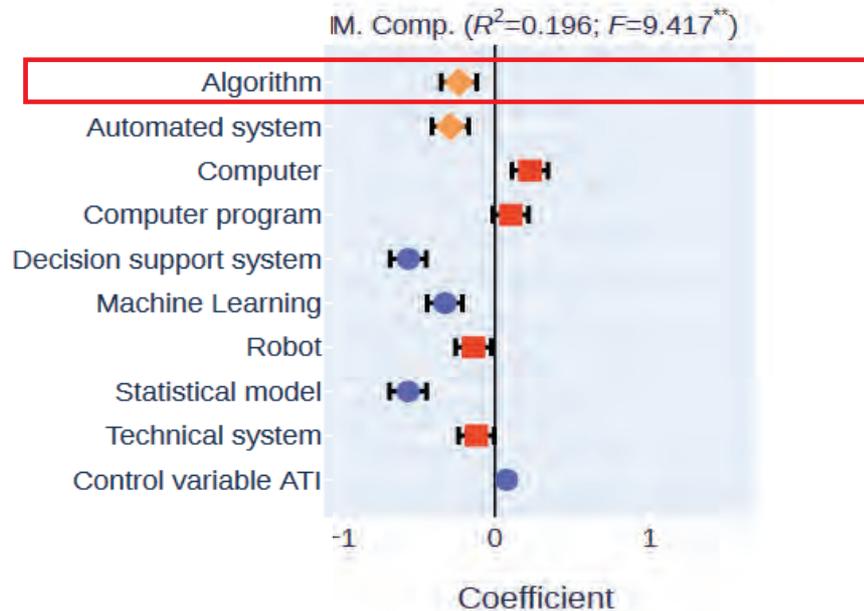
Die neue Macht der Maschinen

Robo Recruiting – Dank Algorithmen bessere Mitarbeiter:innen finden?

## What Hiring Algorithms Mean for Job Searchers Like You

## WORDS MATTER!

LANGER ET AL. (2022)



10. ASSESSMENT KONGRESS 2022

7

## WAS SIND EIGENTLICH DIESE „ALGORITHMEN“?

- Eine Vorgehensweise, um ein Problem zu lösen; ein Rechenvorgang nach einem bestimmten Schema (Duden)
- Algorithmen nehmen Inputs (z.B. Prädiktoren oder Features), verarbeiten diese (z.B. mittels statistischer Verfahren), geben Outputs aus (z.B. Klassifikationen, Vorhersagen, Bewertungen...)
- Beispiel für einen Algorithmus: Body-Mass-Index (BMI)
  - Input: Gewicht + Körpergröße
  - Formel: Gewicht / Körpergröße in Quadrat
  - Output: Body-Mass-Index

10. ASSESSMENT KONGRESS 2022

8

## KLASSISCHER ANSATZ

- „Klassischer“ Ansatz: händisch programmierte Algorithmen (bspw. welche Prädiktoren/Features sind wichtig, wie sind die Beziehungen zwischen Inputs und Outputs?)

## MACHINE LEARNING ANSATZ

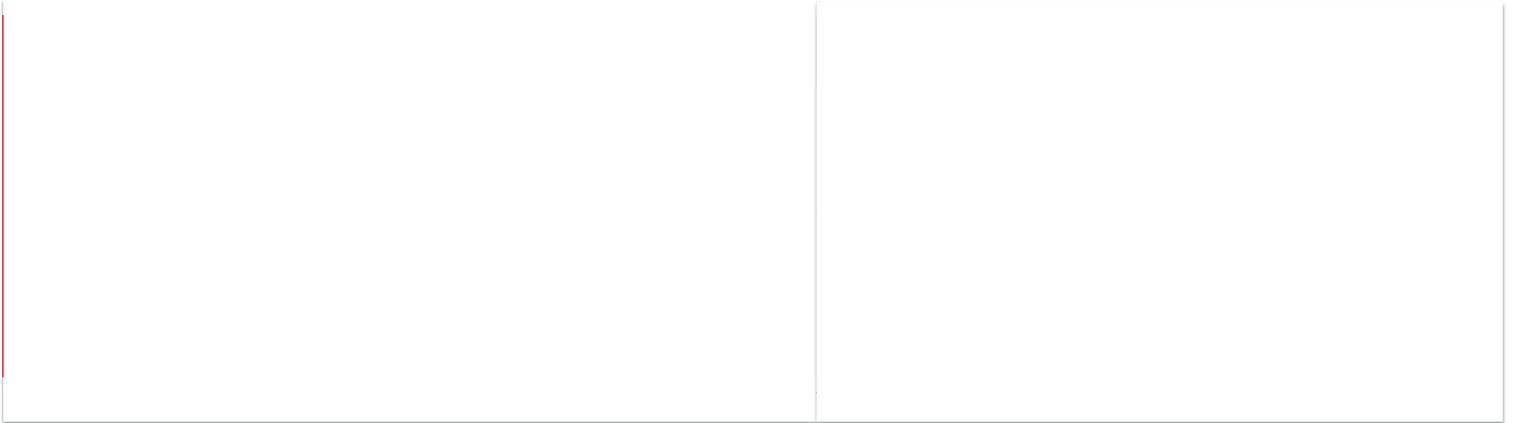
- „Klassischer“ Ansatz: händisch programmierte Algorithmen (bspw. welche Prädiktoren/Features sind wichtig, wie sind die Beziehungen zwischen Inputs und Outputs?)
  - Benötigt: Arbeitszeit, Wissensgrundlage
- Machine Learning Ansatz: Prädiktoren/Features automatisch ausgewählt, Beziehungen zwischen Inputs und Outputs automatisch gelernt
  - Benötigt vor allem viele, hochqualitative Daten

## DO IT YOURSELF!



## VERSCHIEDENE ARTEN VON ALGORITHMEN

- Händisch programmiert
- Selbstlernend
- Kontinuierlich lernend



## **GOLDSTANDARD PROZESS IM SUPERVISED MACHINE LEARNING**

- Wir sind Eisverkäufer, wollen vorhersagen, wie viel Eis wir zum Verkauf in unserem Eiswagen mitnehmen müssen
- Wir haben Daten aus den letzten Jahren vorliegen, die wir nutzen können

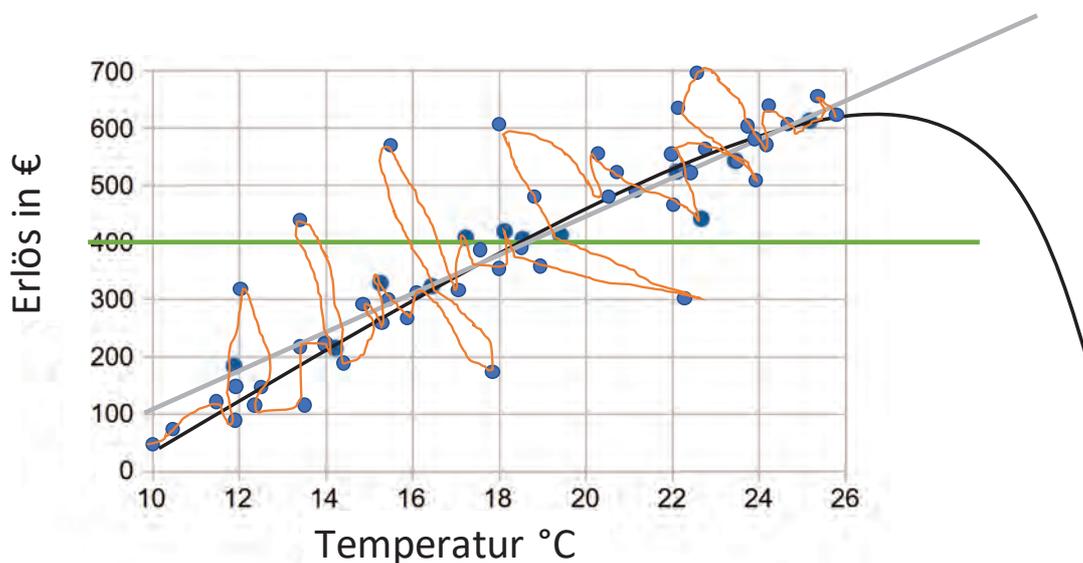
## GOLDSTANDARD PROZESS IM SUPERVISED MACHINE LEARNING



- Wichtig ist, dass diese Schritte in drei getrennten Datensätzen passieren: Trainings-, Validierungs-, Testdatensatz; Beispiel:
  - 80% der vorhandenen Daten fürs Trainieren
  - 20% der vorhanden Daten fürs Validieren
  - Unabhängige, neue Daten fürs Testen
- Denn das Ziel ist: Generalisierbarkeit des Algorithmus

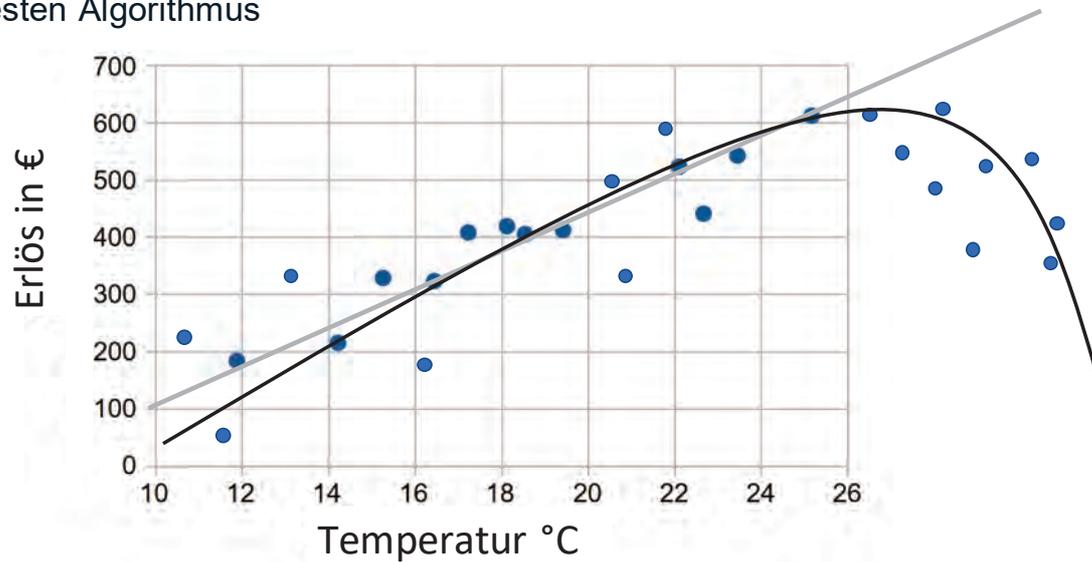
### Trainieren (im Trainingsdatensatz)

- Verschiedene Algorithmen antrainieren



## Validieren (im Validierungsdatensatz)

- Abschätzung der Güte für neue Daten, Auswahl des finalen Modells, Auswahl des besten Algorithmus



## Testen (im Testdatensatz)

- Evaluation der Güte des finalen Modells

## GOLDSTANDARD PROZESS IM SUPERVISED MACHINE LEARNING



**FRAGEN?**

## ABLAUF

- Grundlagen
- Automatisierung in der Eignungsdiagnostik am Beispiel HireVue
- Forschung zu automatisierter Eignungsdiagnostik
- Ausblick

**Hire★Vue**

## UMFRAGE: GRÜNDE DAFÜR, TEILE VON EIGNUNGSDIAGNOSTISCHEN PROZESSEN ZU AUTOMATISIEREN?

1. Kosten und Zeitersparnis
2. Bessere Entscheidungen
3. Bessere "candidate experience"
4. Erhöhung der Fairness/Vermeidung von Biases
5. Andere Gründe

## GRÜNDE FÜR AUTOMATISIERUNG? (LAUT HIREVUE)

- ✓ Biases vermeiden
- ✓ Konsistenz in der Bewertung
- ✓ Chancengleichheit
- ✓ Bessere „candidate experience“
- ✓ Bessere Entscheidungen
- ✓ Kosten und Zeitersparnis

## ENTWICKLUNG DES HIREVUE SYSTEMS

1. Manuelle Bewertung von AVIs – „expert raters using BARS“
2. Automatisiertes Transkribieren von Text aus den AVIs
3. Transformieren des Texts in Zahlen mittels verschiedener Verfahren (z.B. Bag of Words)
4. Trainieren des Machine Learning Modells (Ridge Regression)
5. Validieren des Machine Learning Modells
6. Automatisierte Bewertung von neuen Bewerber:innen
7. Updaten des Modells → d.h. keinen kontinuierlich lernenden Algorithmus genutzt

[https://webapi.hirevue.com/wp-content/uploads/2022/04/HV\\_AI\\_Short-Form\\_Explainability\\_1pager.pdf](https://webapi.hirevue.com/wp-content/uploads/2022/04/HV_AI_Short-Form_Explainability_1pager.pdf)

### WAS PERSONALER\*INNEN SEHEN

**Candidates are displayed by tier**

**Sample candidate list by tiered assessment results (Top, Middle, Bottom)**

**When a candidate is selected a breakdown of how they scored in each competency is displayed (in line with BARS)**

# WAS BEWERBENDE SEHEN

1 2 3 4  
Invited In Progress Processing **Insights Available**

## Your Insights

Thank you for your interest in BB Data and the time you took to complete your responses for Cashier.

We worked with a team of IO Psychologists to develop the interview questions and games you just completed to discover the unique competencies you possess. Scroll through to see your individual results.

<  Working with People  Working Style and Personality  Working with Information >

*"Working with People" describes how you approach building relationships and interacting with others in a work context.*

 **You may find yourself being brief in your interactions with others.**  
*At times, others may need more information and more context, so be sure to check for understanding when speaking with others.*



# FRAGEN?

## ABLAUF

- Grundlagen
- Automatisierung in der Eignungsdiagnostik am Beispiel HireVue
- Forschung zu automatisierter Eignungsdiagnostik
- Ausblick

## FORSCHUNGSSCHWERPUNKTE

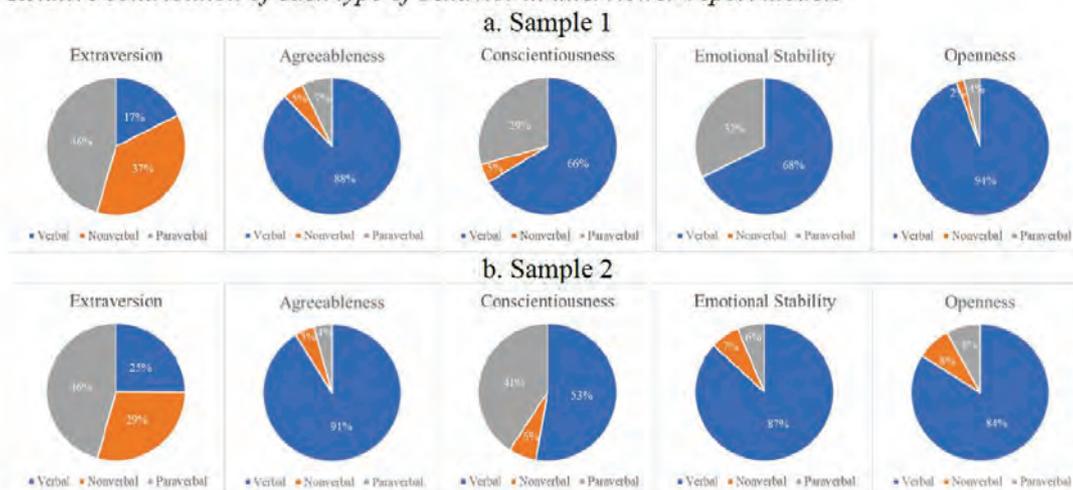
- Validität
- Fairness und Biases
- Akzeptanz

# VALIDITÄT

- Frühe Forschung (2016+) vor allem aus der Informatik
  - Zeigt Potential automatisierter Bewertung für Interviews, z.B. Korrelationen bis zu .62 von automatisierter und menschlicher Bewertung (Naim et al., 2018)
- Champion et al. (2016) als erste ernstzunehmende Studie aus dem Arbeitspsychologischen Bereich zu automatisierter Eignungsdiagnostik (bei Motivationsschreiben)
- Hickman et al. (2021) zeigen in mehreren Studie automatisierte Bewertung von Persönlichkeit aus Interviews
  - Z.B. Korrelationen von .65 zwischen Interviewerbewertungen von Extraversion und automatisierten Bewertungen
  - Zeigen, dass Fremdbewertungen und automatisierte Bewertungen stärker konvergieren als Selbstberichte

## HICKMAN ET AL. (2021)

*Relative contribution of each type of behavior in interviewer-report models*



Blau = Verbal  
 Grau = Paraverbal  
 Orange = Nonverbal

(Meisten Provider nutze nur noch verbale Inhalte)

## HERAUSFORDERUNGEN VALIDITÄT

- Keine „echten“ Assessmentprozesse
- Fraglich, ob es reicht menschliche Bewertungen gut nachzubilden – braucht es andere Arten von Kriterien?
- Provider (wie HireVue) behaupten, sie haben genug Validitätsevidenz – die ist aber nicht in wissenschaftlichen Studien verfügbar

## FAIRNESS UND BIASES

Vendor	Claim about bias
HireVue	Provide "a highly valid, bias-mitigated assessment"
pymetrics	"... the Pre-Hire assessment does not show bias against women or minority respondents."
PredictiveHire	"AI bias is testable, hence fixable."
Knockri	"Knockri's A.I. is unbiased because of its full spectrum database that ensures there's no benchmark of what the 'ideal candidate' looks like."

Table 2: Examples of claims that vendors make about bias, taken from their websites.

Häufiges Argument der Provider: Automatisierte Bewertung kann die Fairness des Auswahlprozesses erhöhen durch...

- ...die Verringerung des Einflusses menschlicher Biases
- ...konsistente, weil algorithmenbasierte, Bewertung von Bewerbenden

## HERAUSFORDERUNGEN FAIRNESS UND BIASES

- Sehr schnell können Machine Learning basierte Systeme „einfache“ Entscheidungsregeln lernen („sortiere alle Menschen Ü40 aus“)
- Fairness und Biases als komplexe Probleme (siehe z.B. Tay et al., 2021), die schwer in Zahlen auszudrücken sind; nicht alle „Fairnessregeln“ sind gleichzeitig erfüllbar
- Positiv anzumerken: die Automatisierung von Prozessen ist ein super Anreiz, um strukturiert über Fairness und Biases nachzudenken

## AKZEPTANZ

- Auf Seite der Personen, die automatisiert bewertet werden, gibt es einiges an Forschung (zusammengefasst z.B. von Langer & Landers, 2021)
- Besonders für Interviews wird die automatisierte Bewertung von Bewerbenden als kritisch angesehen
  - Obwohl erwartet wird, dass Automatisierung Konsistenz bringt und Biases verringern kann
- Eher akzeptiert wenn:
  - Frühere Phasen des Bewerbungsprozesses bzw. weniger „Erwartung“ menschlichen Austausches
  - Menschliche Aufsicht bzw. letzte Entscheidung bei einem Menschen

## AKZEPTANZ

- Auf Seite der Anwender\*innen weniger Forschung
- Contra (Li et al., 2021)
  - High-Potentials haben keine Lust auf automatisierte Bewertung
  - Rolle der Anwender\*innen verändert sich, z.B. muss man Akkuratheit von Systemen prüfen (Langer et al., 2022)
  - Systemoutputs wenig nachvollziehbar
- Pro (Li et al., 2021)
  - Effizienzgewinn
  - Ansprache eines weiteren Bewerbendenpools
  - Evtl. weniger Biases

## HERAUSFORDERUNGEN AKZEPTANZ

- Keine „echten“ Assessmentprozesse
- Es gibt nicht „DAS“ System – Systemdesignentscheidungen bisher wenig beachtet
- Erwartungen an Systeme führen möglicherweise zu problematischem Verhalten (siehe z.B. Langer et al., 2022)
  - Wenn Provider behaupten, Systeme sind weniger biased, Bewerbende und Anwender\*innen das erwarten, dann fallen Biases in Systemoutputs möglicherweise weniger auf

## GRUNDLEGENDE: GÜTE AUTOMATISIERTER VERFAHREN?

- Wie können Unternehmen die Güte automatisierter Verfahren bewerten?
- Rückfrage: Wie bewerten Unternehmen die Güte von Persönlichkeitstests? Von Intelligenztests?
- → Eigentlich verändert sich nicht viel!
  - Spezifische Liste an Fragen siehe nächste Folie

## FRAGEN ZUR SYSTEMENTWICKLUNG

LANDERS & BEHREND (2022)

- Fragen zu Input Daten, Model Design, Model Development, Model Features, Model Prozesse, Model Outputs
- Fragen zu Reaktionen von Bewerbenden, Anwendenden, der Öffentlichkeit
- Fragen zum Einfluss des kulturellen Kontexts, nationaler Rechtsprechung
- Folgt die Systementwicklung und Implementierung (ethischen und kontextspezifischen) Standards? (z.B. DIN 33430)
- Gibt es Forschung und wie sieht diese Forschung aus?

## ABLAUF

- Grundlagen
- Automatisierung in der Eignungsdiagnostik am Beispiel HireVue
- Forschung zu automatisierter Eignungsdiagnostik
- Ausblick

## AUSBLICK (I)

- Unternehmen haben Daten, implementieren Systeme, setzen Systeme in echten Prozessen ein... all das ist zentral für die Forschung
- Unternehmen haben jedoch womöglich wenig Anreiz Forschung in Journals zu veröffentlichen, auch Datenschutzthemen mögen eine Herausforderung darstellen
- Dennoch → Kooperationen mit Unternehmen zentral für die zukünftige Forschung

## AUSBLICK (II)

- Aktuelle Entwicklungen im Rechtsrahmen des Einsatzes automatisierter Systeme – besonders der AI Act der Europäischen Kommission
- Jegliche KI-basierten Systeme, die im Personalwesen eingesetzt werden, werden als „Hochrisikosysteme“ eingestuft
- Damit ergeben sich besondere Anforderungen an diese Systeme; besonders betont im AI Act: menschliche Aufsicht, Transparenz und Nachvollziehbarkeit

## TAKE HOME

- Automatisierung in der Eignungsdiagnostik keine Magie, eher angewandte Statistik, psychometrische Gütekriterien bleiben wichtig
  - Anforderungen an „KI“ im Assessment im Großen und Ganzen gleichzusetzen mit Anforderungen an klassische Persönlichkeitstests
  - Dennoch teils „magische“ Erwartungen
- Kooperationen zwischen Forschung und Praxis dringend nötig, um sinnvolle, forschungsbasierte Empfehlungen geben zu können
- Mit Spannung erwartet werden die Implikationen aktueller Vorschläge zu Rechtsvorschriften im Bereich KI

## FRAGEN?



Vielen Dank

für Ihre Aufmerksamkeit

[Markus.langer@uni-saarland.de](mailto:Markus.langer@uni-saarland.de)

Twitter: @LangersMarkus

### **Tummelplatz für Blender oder fundierter Erkenntnisgewinn – Was bringen ACs wirklich für die Personalauswahl?**

Prof. Klaus Melchers & Tina Uhle

Assessment Center (ACs) sind ein vergleichsweise teures und aufwändiges Verfahren der Personalauswahl. Zwar sind sie etabliert und weitgehend akzeptiert, gleichzeitig gibt es sowohl aus der Wissenschaft als auch aus der Praxis immer wieder Kritik an ihnen. Ziel dieses Beitrags ist einerseits zu beleuchten, wie stichhaltig geäußerte Kritikpunkte bzgl. ACs tatsächlich sind und wie man mit ihnen umgehen kann. Zum anderen geht es darum, was beachtet werden muss, damit ACs als Auswahlverfahren möglichst gut funktionieren.

#### **Kritiker äußern: Im Vergleich zu anderen Auswahlverfahren ist die Kriteriumsvalidität von ACs nur mäßig hoch!**

Mit Bezug auf die bekannte Überblicksarbeit von Schmidt und Hunter (1998) erscheinen ACs im Vergleich zu anderen Verfahren nur eine mittelmäßige Kriteriumsvalidität zu besitzen. Bzgl. dieses Überblicks sind jedoch einige wichtige Dinge zu beachten: Erstens gibt es mittlerweile eine aktuellere metaanalytische Übersicht von Sackett et al. (2021), die zeigt, dass ACs im Vergleich zu vielen anderen Verfahren eine gute Kriteriumsvalidität haben. Zweitens werden bei Metaanalysen zu Kriteriumsvaliditäten jeweils Korrekturen für die Unreliabilität der Kriteriumsmessung (d. h. der Arbeitsleistung) und der Varianzeinschränkung im Auswahlverfahren vorgenommen. Insbesondere letztere führen zu deutlichen Erhöhungen der Schätzungen der „wahren“ Validität. Bei ACs sind diese Korrekturen im Vergleich zu anderen Verfahren wie z. B. IQ-Tests jedoch kaum möglich. Und drittens gibt es eine Reihe von Ergebnissen, die belegen, dass Aspekte, die zur Qualität von ACs beitragen, sich positiv auf die Kriteriumsvalidität auswirken.

#### **Kritikpunkt: ACs wurden über die Zeit immer schlechter!**

Neuere metaanalytische Schätzungen der Kriteriumsvalidität von ACs fallen niedriger aus als ältere. Dies trägt zur Befürchtung bei, dass ACs immer schlechter wurden. Mögliche Gründe, warum Validitäten heutzutage niedriger ausfallen liegen jedoch auch darin, dass die Vorauswahl heutzutage oftmals durch billigere, kriteriumsvalide Verfahren wie IQ-Tests erfolgt und Validitätsstudien häufig mit Stelleninhaber\*innen statt Bewerber\*innen durchgeführt werden. Beides führt dazu, dass die Varianz der AC-Leistung geringer ist und die gefundenen empirischen Validitäten deswegen kleiner ausfallen. Diese Varianzeinschränkung ist in den entsprechenden Metaanalysen jedoch nicht einschätzbar und korrigierbar. Zudem finden sich in manchen Metaanalysen trotz expliziter Testung keine Belege, dass neuere ACs tatsächlich schlechter sind als ältere.

## 10. Assessment Kongress

### **Kritik: Wenn man bereits einen Intelligenztest verwendet, bringt ein zusätzliches AC (fast) nichts mehr! Und sind Interviews nicht sowieso immer besser?**

Ebenfalls auf Basis der Übersichtsarbeit von Schmidt und Hunter (1998) wird regelmäßig der Einwand erhoben, dass die inkrementelle Validität (d. h. die Verbesserung der Vorhersageleistung) bei Verwendung eines ACs zusätzlich zu einem IQ-Test vernachlässigbar gering ist. Zudem haben standardisierte, anforderungsbezogene Interviews tatsächlich eine bessere Kriteriumsvalidität als ACs. In Bezug auf Schmidt und Hunter (1998) gibt es jedoch methodische Probleme, die dafür sprechen, dass die inkrementelle Validität von ACs dort unterschätzt wurde. Neuere metaanalytische Ergebnisse deuten ebenfalls darauf hin, dass ACs auch bei vorheriger Verwendung von IQ-Tests zu einer deutlichen Steigerung der Validität beitragen können – und bei hoch qualifizierten Bewerbendenstichproben sogar valider als IQ-Tests sein können. Und in Bezug auf Interviews haben ACs den Vorteil, dass sie echtes Verhalten erfassen und deutlich weniger von möglichem Faking betroffen sind als Interviews. Zudem sind Interviews in der Praxis oftmals nicht so standardisiert und anforderungsbezogen, wie dies aus eignungsdiagnostischer Sicht empfehlenswert ist – was die Konsequenz hat, dass ihre mögliche Kriteriumsvalidität nicht maximal ausgeschöpft wird.

### **Kritischer Einwand: ACs sind nicht konstruktvalide, deswegen sind sie auch als Auswahlverfahren nicht brauchbar und nicht vertrauenswürdig!**

Beurteilungen aus ACs bilden tatsächlich vor allem Leistungsunterschiede zwischen verschiedenen Übungen ab, jedoch weniger zwischen verschiedenen Anforderungsdimensionen. Dies führte zu der lange Zeit dominierenden Sicht, dass sich im AC vor allem Übungseffekte anstatt Konstrukte (also Anforderungsdimensionen oder Kriterien) zeigen und dies ein Beobachtendenfehler ist. Auch Maßnahmen zur Verbesserung der Konstruktvalidität (z. B. Training für Beobachtende, eine Reduktion der zu beobachtenden Dimensionen oder die Verwendung von Verhaltenschecklisten) verbessern die Konstruktvalidität in der Regel nur in geringem Maß (obwohl sie sich allerdings z. T. positiv auf die Kriteriumsvalidität auswirken). Allerdings sprechen eine Reihe von Ergebnissen dafür, dass hohe Korrelationen zwischen Leistungsbeurteilungen innerhalb der einzelnen AC-Übungen nicht nur durch mögliche Fehler der Beobachtenden zustande kommen. Ganz im Gegensatz ist es eher so, dass diese hohen Korrelationen durch tatsächliche Leistungsunterschiede innerhalb der Bewerber\*innen zwischen unterschiedlichen Situationen herrühren – und wenn die Situationen (d. h. die AC-Übungen) sinnvoll ausgewählt wurden, trägt dies sogar wesentlich zur Kriteriumsvalidität der ACs bei.

### **Kritiker postulieren: Im AC wird vor allem Sympathie beurteilt! Und ACs sind Verfahren, die vor allem extravertierte Vielredner\*innen bevorzugen und die besten Industrieschauspieler\*innen belohnen!**

Tatsächlich gibt es einige im AC bedeutsame Beobachtungs- und Bewertungsfehler (z. B. Verzerrungen durch Vorinformationen, Halo-Effekte, Reihenfolgeeffekte, ...). Deren Einfluss kann jedoch minimiert werden durch eine sorgfältige Schulung der Beobachtenden

## 10. Assessment Kongress

sowie die Einhaltung bestimmter Qualitätsregeln. Dazu gehören bspw. die Pflicht zum detaillierten Mitprotokollieren, der Einsatz einer geschulten Moderation, die Nutzung von Entscheidungsregeln und einiges mehr. Auch diese Regeln sollten im Training der Beobachter\*innen vermittelt und eingeübt werden. Um die Akkuratheit der Beobachtungen und Beurteilungen zu verbessern, sind insbesondere Verhaltensbeobachtungs- und Bezugsrahmentrainings empfehlenswert. Und im Hinblick auf die Entscheidungsregeln ist zentral, dass im Vorfeld sinnvolle Regeln entwickelt werden, aus denen hervorgeht, wie die im AC gewonnenen Informationen in die Beurteilungen einfließen und wie einzelne Bewertungen kombiniert werden. Durch eine konsistente Anwendung der Regeln kann sichergestellt werden, dass die verschiedenen validen Einzelinformationen tatsächlich optimal berücksichtigt werden.

### **Kritik: Das AC produziert Verlierer\*innen und ist ein Motivationskiller**

ACs wird auch vorgeworfen, dass sie Verlierer\*innen produzieren und deswegen ein Motivationskiller sind. Dies ist allerdings kein AC-spezifisches Problem, sondern ein Effekt, der bei jeder Art von Karriereentscheidung ansteht, unabhängig von der Methode AC. Bei einem gut gemachten AC ist der Entscheidungsprozess jedoch transparent, so dass Teilnehmenden nachvollziehen können, wie eine Entscheidung zustande gekommen ist. Zudem bietet ein gutes AC Lernchancen, die sich aus einem differenzierten Feedback an die Teilnehmenden ergeben oder die Rückmeldung aus Beteiligtersicht, dass ein AC durchaus Spaß machen und ein anspruchsvoller Auswahlprozess auch für den Qualitätsanspruch eines Unternehmens sprechen kann.

### **Fazit**

- Gut entwickelte ACs gehören eindeutig zu den Auswahlverfahren mit guter Validität
- ACs wurden nicht im Lauf der Jahre schlechter. Stattdessen wurde vermutlich vor allem die Vorauswahl besser bzw. strenger
- Intelligenztests und (strukturierte) Interviews sind gute Auswahlverfahren – ACs bieten dennoch einen Mehrwert
- Die oft bemängelte geringe Konstruktvalidität kommt nicht nur durch Fehler der Beobachtenden sondern auch durch valide situationsspezifische Leistungsunterschiede zustande
- Wahrnehmungsverzerrungen können durch hochwertig und standardisiert gestaltete ACs und Beobachtendenschulungen minimiert werden

Die im Jahr 2016 vom Forum Assessment aktualisierten „Standards der Assessment Center Methode“ stellen konkrete Ansatzpunkte für eine qualitativ hochwertige Entwicklung und Durchführung von ACs dar. ACs, die die dort formulierten Standards einhalten, vermeiden die verschiedenen geäußerten Kritikpunkte und können einen zentralen Beitrag im Rahmen der Personalauswahl leisten.



BEITRAG 14-AC

## TUMMELPLATZ FÜR BLENDER ODER FUNDIERTER ERKENNTNISGEWINN

– WAS BRINGEN ACs WIRKLICH FÜR DIE PERSONALAUSWAHL?  
PROF. DR. KLAUS MELCHERS & TINA UHLE

## REFERENTINNEN



Prof. Dr. Klaus Melchers

- Professor für Arbeits- & Organisationspsychologie an der Universität Ulm
- knapp 20 Jahre Forschung zu eignungsdiagnostischen Fragestellungen
- Mitglied des Wissenschaftlichen Beirats des Forum Assessment



Tina Uhle

- Geschäftsführerin Uhle Unternehmensberatung GmbH
- 20 Jahre praktische Entwicklung und Durchführung eignungsdiagnostischer Instrumente für unterschiedliche Zielgruppen
- Engagiert sich im Verein seit 2004

## WARUM WIR HIER SIND... ODER: WIE ALLES BEGANN...

Wissenschaft und Praxis

Forschung und Umsetzung

Erkenntnisse und  
Standards

Vorlesung und Firma



Fundierung und Erfahrung

Klaus und Tina

## ERGEBNISSE MENTIMETER: HÄUFIG GEÄUSSERTE RITIKPUNKTE AN ACs



## Im Vergleich zu anderen Auswahlverfahren ist die Kriteriumsvalidität von ACs nur mäßig hoch!

## EINERSEITS: ÜBERSICHT VON SCHMIDT & HUNTER (1998)

Metaanalytische Kriteriumsvaliditäten für die Vorhersage von Arbeitsleistung:

