



VALIDIZACE VÝBĚROVÝCH ŘÍZENÍ: MOŽNOSTI VYUŽITÍ HIERARCHICKÝCH LINEÁRNÍCH MODELŮ

Patrícia Martinková

Ústav informatiky AV ČR

Dan Goldhaber

Center for Education Data & Research
University of Washington, Bothel

Projekty:

- Fulbright-Masarykovo stipendium, 2013 – 14
 - Center for Statistics and the Social Sciences, University of Washington, USA
 - projekt „*Generalizing psychometric frameworks for estimating the quality of psychological, educational and health measurements*“
- Projekt financovaný IES (Institute of Education Sciences)
 - „*Applicants at the Doorstep: Improving Hiring Practices through a Better Understanding of the Link between Applicant Information and Teacher Quality*“
 - Researcher – Practitioner grant (CEDR a školský distrikt Spokane, WA, USA)
- Juniorský projekt GA ČR 2015 – 2017
 - „*Odhad psychometrických vlastností jako součást vývoje přijímacích testů*“

- **Publikace**

- Goldhaber, Grout & Huntington-Klein: *Screen Twice, Cut Once: Assessing the Predictive **Validity** of Teacher Selection Tools*. CEDR WP #2014-9
- Martinková & Goldhaber: *Improving Teacher Selection: The Effect of Intra-Rater **Reliability** in the Screening Process*. CEDR WP #2015-7
- Martinková & Vlčková: *Hodnocení reliability znalostních a psychologických testů*. Informační bulletin České statistické společnosti, 25(4), 2014, str. 1-15.

- **Prezentace**

- Association of Educational Finance and Policy (AEFP 2015), Washington D.C.
- Joint Statistical Meetings (JSM 2015), Seattle

Úvod: Důležitost přijímacích řízení

Přijímací řízení je klíčovým momentem pro uchazeče i instituci:

- Kvalita uchazeče ovlivňuje kvalitu školy a naopak
- Stanovením kritérií pro přijetí vyvíjí škola tlak na kvalitu a přípravu uchazečů

Úvod: Zajištění efektivity přijímacích řízení

4

- Zajištění co největšího počtu kvalitních uchazečů
- Vysoká **validita** hodnocení
 - Hodnocení dobře predikuje budoucí kvality uchazeče
- Vysoká **reliabilita** hodnocení
 - Hodnocení stejného uchazeče pro stejný typ pozice jsou konzistentní
- Strukturovanost
- Dobře stanovená a popsaná kritéria hodnocení
- Vícestupňové řízení (při vysokém počtu uchazečů)

Úvod: Cíle prezentace (osnova)

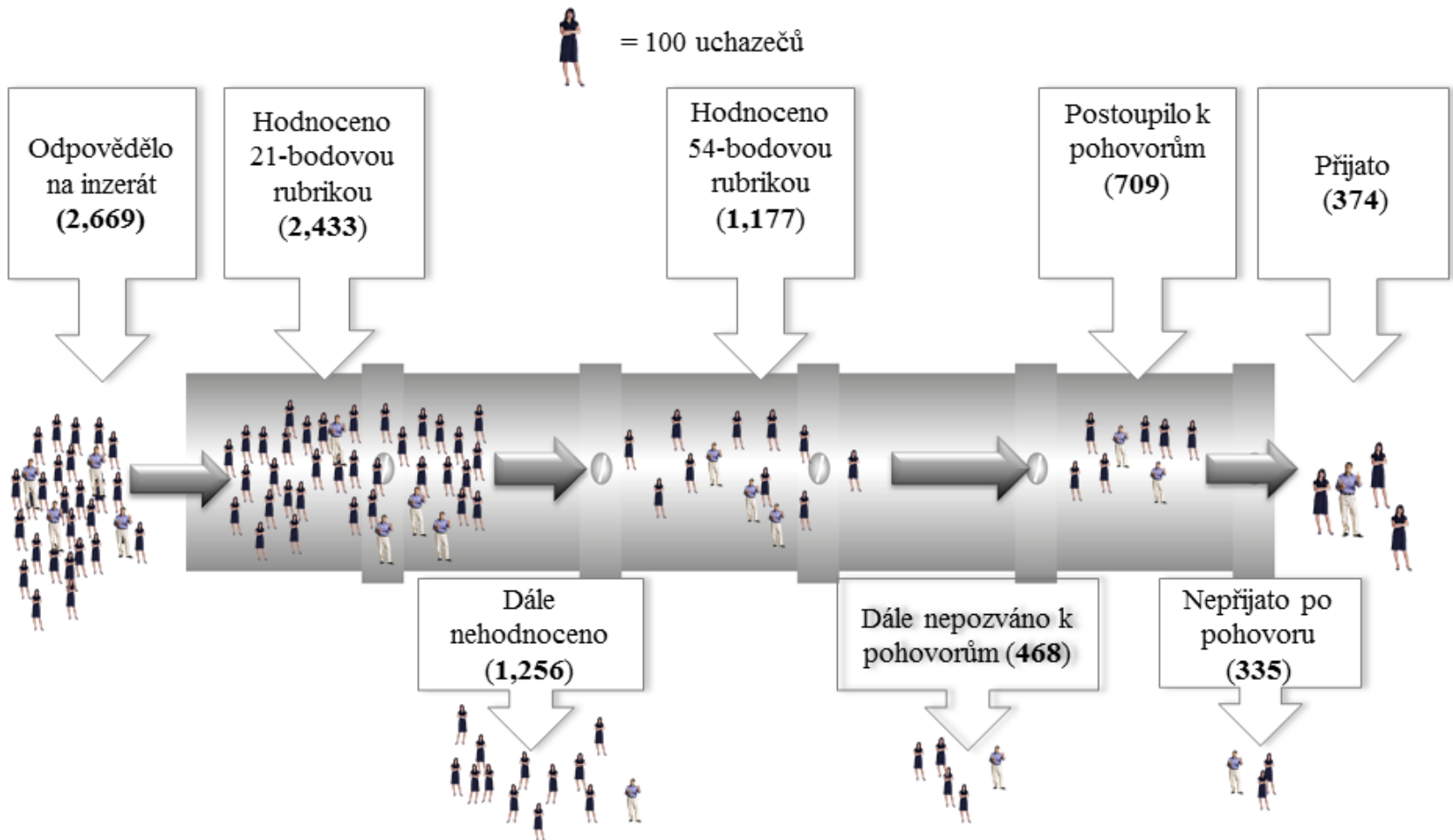
5

- Případová studie: strukturované vícestupňové výběrové řízení na ZŠ/SŠ v oblasti Spokane, USA
- Prokazování **validity** hodnocení uchazečů
 - Především s ohledem na přidanou hodnotu učitele (teacher value added)
- Odhad **inter-rater reliability** hodnocení ve složitějších designech
 - Velké množství hodnotitelů
 - Design není vyvážený
 - Velké množství kovariát, které mohou ovlivnit reliabilitu hodnocení
 - Typ školy (ZŠ/SŠ)
 - Typ uchazeče (interní/externí)
 - Typ hodnotitele (zkušený/nezkušený)
 - Vliv reliability na validitu hodnocení
- Diskuse: možnosti uplatnění prezentovaných metod v ČR

Výběr učitelů - Spokane, USA

Uchazeči

7



Uchazeči o učitelské pozice v školském distriktu Spokane, 2008/09 – 2012/13

54-bodová rubrika

8

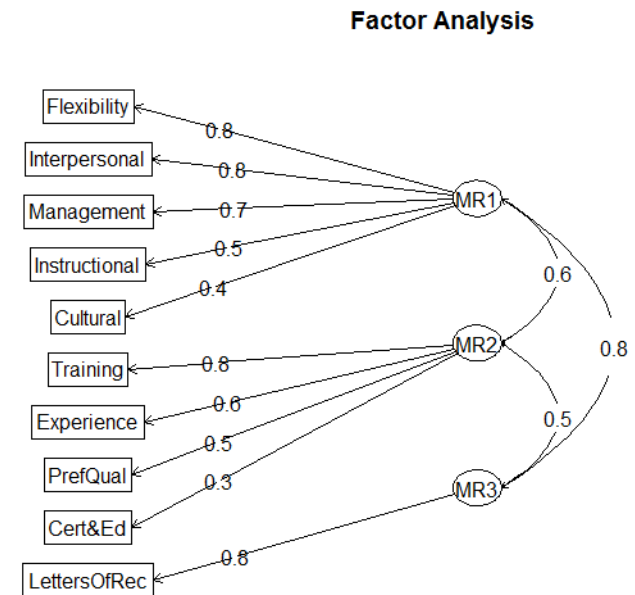
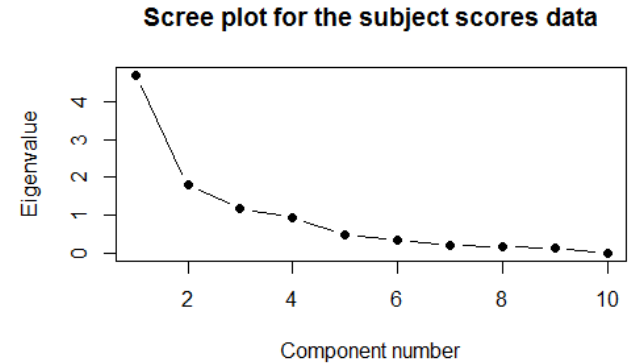
- Vzdělání a certifikáty
 - Praxe (Training)
 - Zkušenosti (Experience)
 - Vedení třídy
(Classroom management)
 - Flexibilita
 - Instruktažní dovednosti
 - Interpersonální dovednosti
 - Kulturní kompetence
 - Preferované kvalifikace
-
- Pozn.: V roce 2013 se objevuje navíc kritérium Kvalita doporučujících dopisů

CERTIFICATED APPLICANT - PRINCIPAL / SUPERVISOR SCREENING	
DATE:	SCREENER:
Job # / Position Title:	
APPLICANT NAME:	
SCREENING CRITERIA	RATING (1-6) 5 - 6 Strong evidence to support this as an area of strength 3 - 4 Satisfactory evidence to support this as an area of strength 1 - 3 Some evidence to support this as an area of strength
CERTIFICATE AND EDUCATION	Note completion of course of study, certificate held (current or pending); education
Washington State Certificate	Yes / No
Required Endorsement	Yes / No
Rating (1 - 6)	4
TRAINING	Look for quality, depth and level of candidates additional training relating to the position.
Rating (1 - 6)	4
EXPERIENCE	Note degree to which experience supports the prediction of success ... not just the number of years. A beginning candidate could be rated highly
Rating (1 - 6)	4
CLASSROOM MANAGEMENT	Look for specific references to successful strategies. This may not mean quiet and orderly but planned and directed. Effectively handles large, small or ethnically/socioeconomic ally diverse groups; develops routines and procedures to increase learning, establishes clear parameters, and responds appropriately.
Rating (1 - 6)	4
FLEXIBILITY	Note multiple endorsements, activity, coaching interests, student, building or district, or community support. Willing to learn new concepts and procedures, successfully teaches a variety of assignments, effectively uses various teaching styles.
Rating (1 - 6)	4
INSTRUCTIONAL SKILLS	Look for specific references in support of skill in this area - plans, implements, evaluates, relates to students, creative, multiple approaches, monitors and adjusts, uses culturally responsive strategies appropriate to age, background and intended learning of students.
Rating (1 - 6)	4
INTERPERSONAL SKILLS	Develops and maintains effective working relationships with diverse staff, students, parents/guardians, and community.
Rating (1 - 6)	4
CULTURAL COMPETENCY	Look for specific references to successful strategies for building and maintaining a relationship with each student and their family. This may not be explicitly mentioned, but the following strategies offer some evidence of cultural competency: specific instructional strategies providing each student access to a rigorous curriculum, inclusive/respectful language about students and families, a belief that all children can achieve at high levels, mention of conflict resolution/restorative practices, specific instructional strategies for integrating culturally responsive materials which are also rigorous, and appropriate statements about their work with diverse populations. Note relevant training, course work, authors/book titles listed.
A competency based on the premise of respect for individual and cultural differences (race, religion, sexual orientation, gender, abilities, socio-economic status, etc.) and regular implementation of a trust-promoting inclusion.	
Rating (1 - 6)	4
PREFERRED QUALIFICATIONS AS INDICATED ON POSTING	
Rating (1 - 6)	4
LETTERS OF RECOMMENDATION	Look for current letters of recommendation from most the most recent supervisor(s). Your score should reflect the quality and recency of the recommendation as well as the author of the letter. (Example: Are the letters from peers or current supervisors?)
Rating (1 - 6)	4
TOTAL SCREENING SCORE	40

Korelace, faktorová analýza

9

	Cert&Ed	Training	Experience	Management	Flexibility	Instructional	Interpersonal	Cultural	PrefQual	LettersOfRec
Cert&Ed	1	0.25	0.27	0.23	0.24	0.23	0.26	0.19	0.31	0.21
Training	0.25	1	0.59	0.43	0.44	0.52	0.41	0.41	0.46	0.39
Experience	0.27	0.59	1	0.51	0.5	0.57	0.48	0.42	0.48	0.5
Management	0.23	0.43	0.51	1	0.69	0.73	0.66	0.48	0.42	0.69
Flexibility	0.24	0.44	0.5	0.69	1	0.68	0.72	0.51	0.42	0.66
Instructional	0.23	0.52	0.57	0.73	0.68	1	0.68	0.5	0.5	0.72
Interpersonal	0.26	0.41	0.48	0.66	0.72	0.68	1	0.53	0.42	0.65
Cultural	0.19	0.41	0.42	0.48	0.51	0.5	0.53	1	0.37	0.49
PrefQual	0.31	0.46	0.48	0.42	0.42	0.5	0.42	0.37	1	0.56
LettersOfRec	0.21	0.39	0.5	0.69	0.66	0.72	0.65	0.49	0.56	1



Validita výběrového řízení

Goldhaber, Grout & Huntington-Klein: *Screen Twice, Cut Once: Assessing the Predictive **Validity** of Teacher Selection Tools.*
CEDR WP #2014-9

Predikční validita výběrového řízení

11

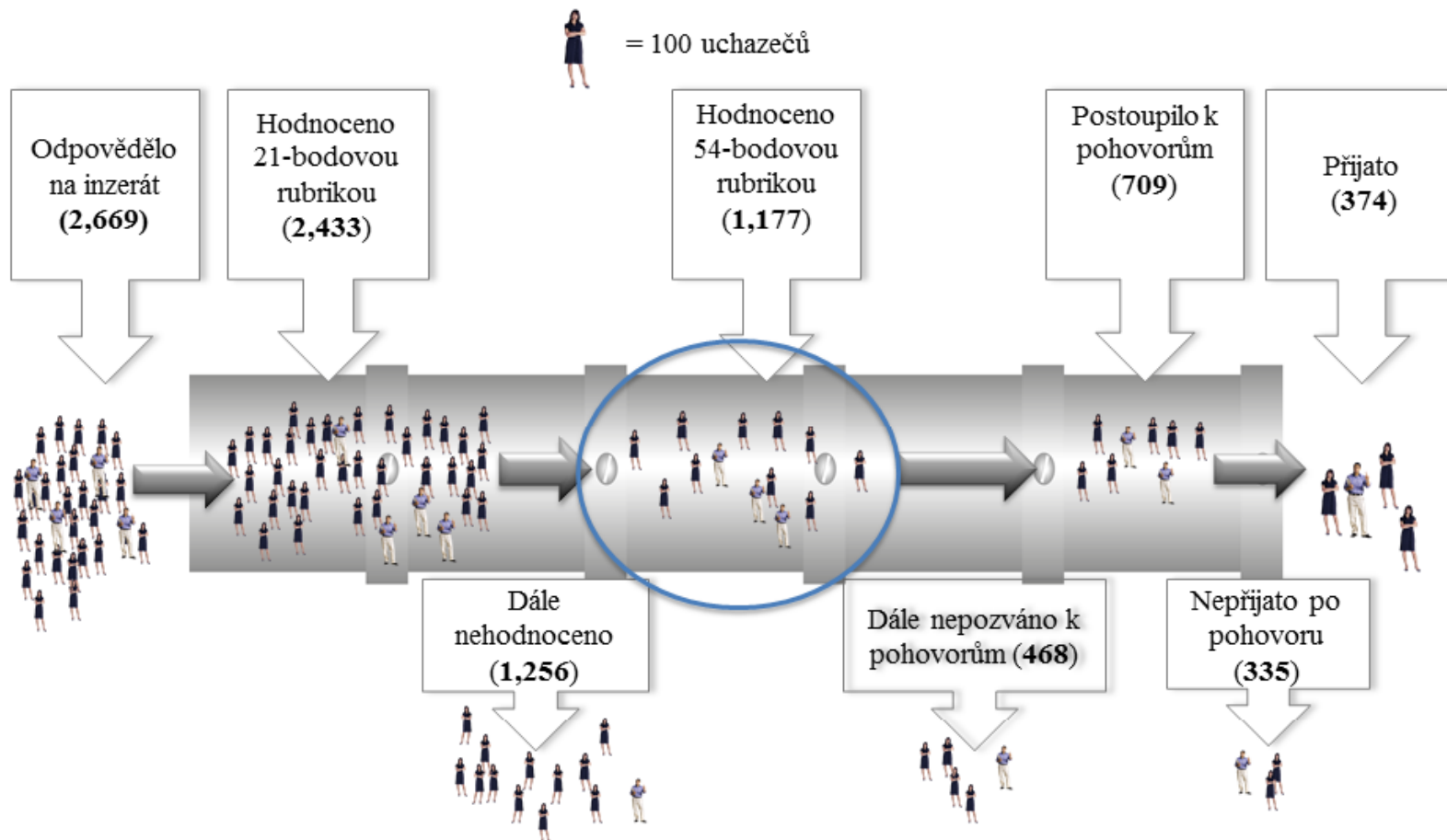
- Goldhaber et al. (2014) ukazují, že skóre v rubrikách je asociováno s
 1. Přidanou hodnotou (Value-added) učitele
 2. Pravděpodobností změny školy
- Z jednotlivých kritérií jsou prediktivní obzvláště:
 - Vedení třídy, Flexibilita, Instruktažní dovednosti, Výcvik
- Analýzy zahrnují uchazeče/učitele, kteří nebyli přijati v distriktu Spokane, ale byli přijati v jiném distriktu ve státě Washington
- Analýzy se vypořádávají s možným výběrovým zkreslením (selection bias) pomocí tzv. instrumentální proměnné
 - Chyby v součtech použity jako instrumentální proměnná

Inter-Rater Reliability

Martinková & Goldhaber: *Improving Teacher Selection:
The Effect of Intra-Rater **Reliability** in the Screening Process.*
CEDR WP #2015-7

Hodnocení Inter-Rater Reliability: Data

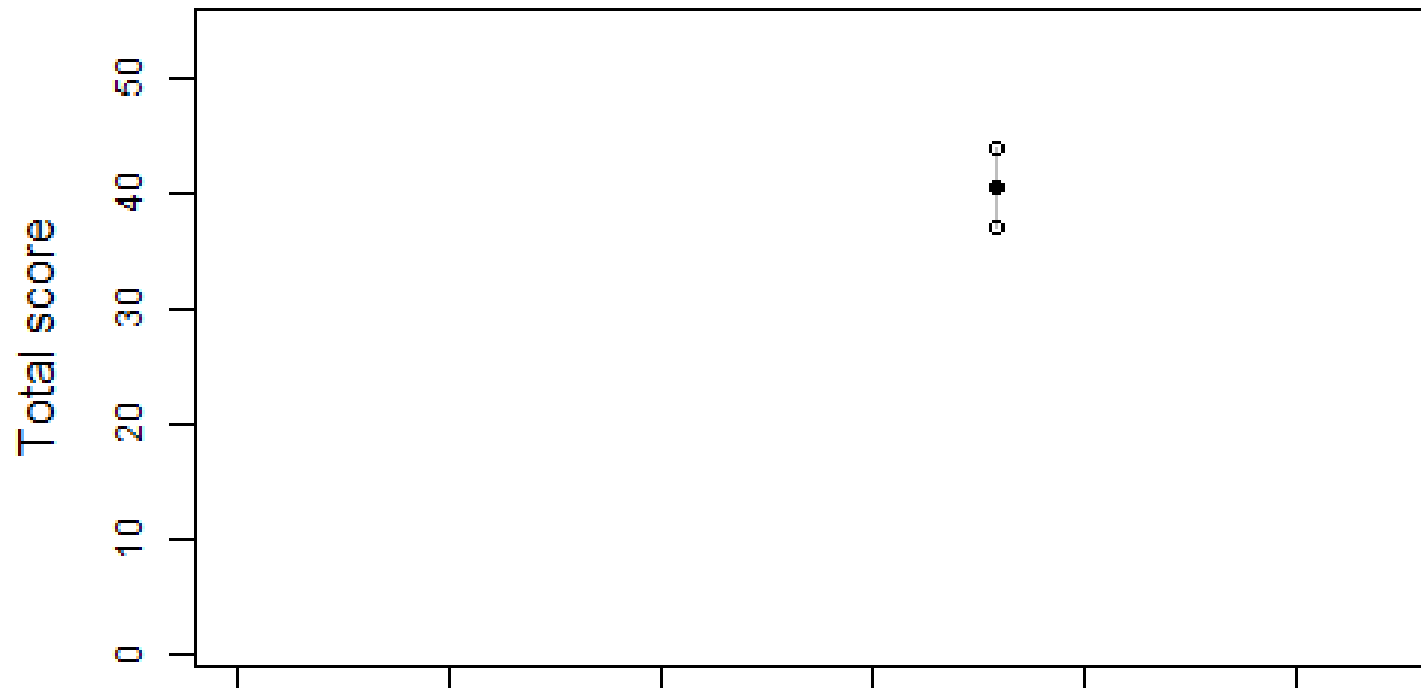
13



Motivace: (Ne)konzistence hodnocení

14

Mean and range of ratings

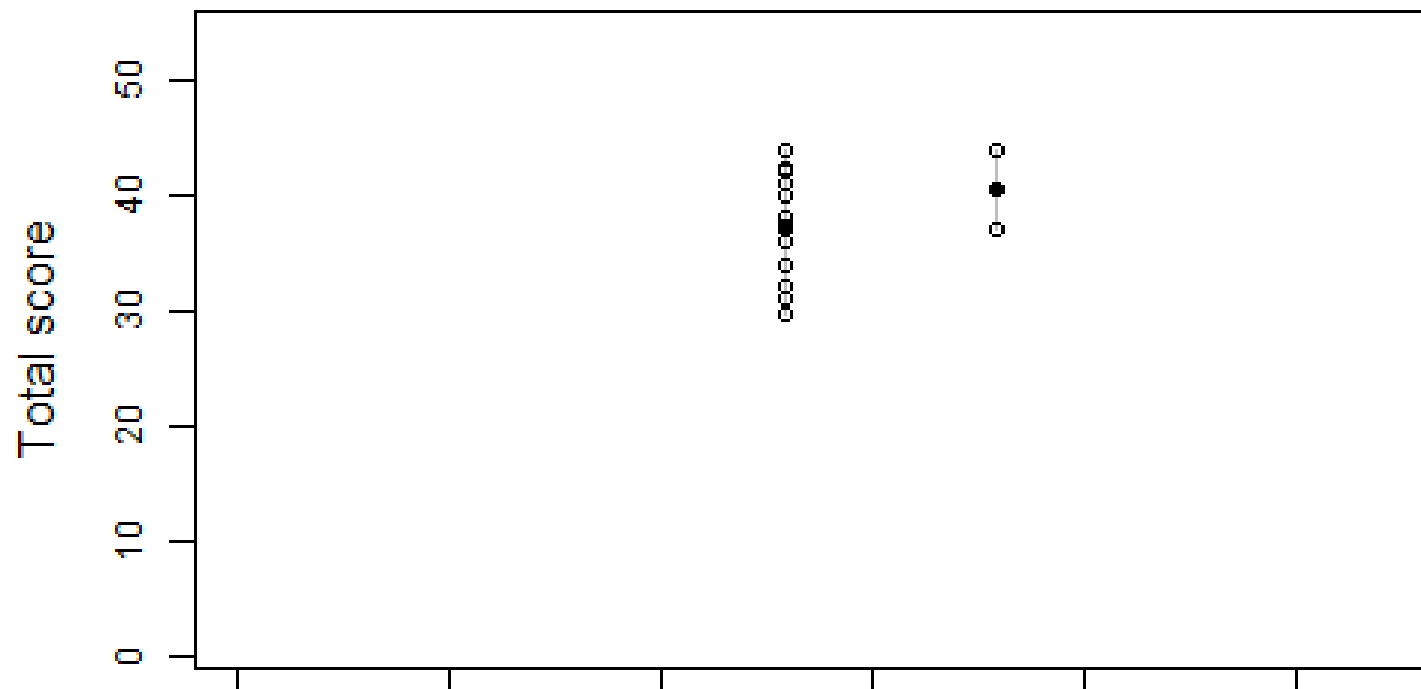


Applications ranked by average total score

Motivace: (Ne)konzistence hodnocení

15

Mean and range of ratings

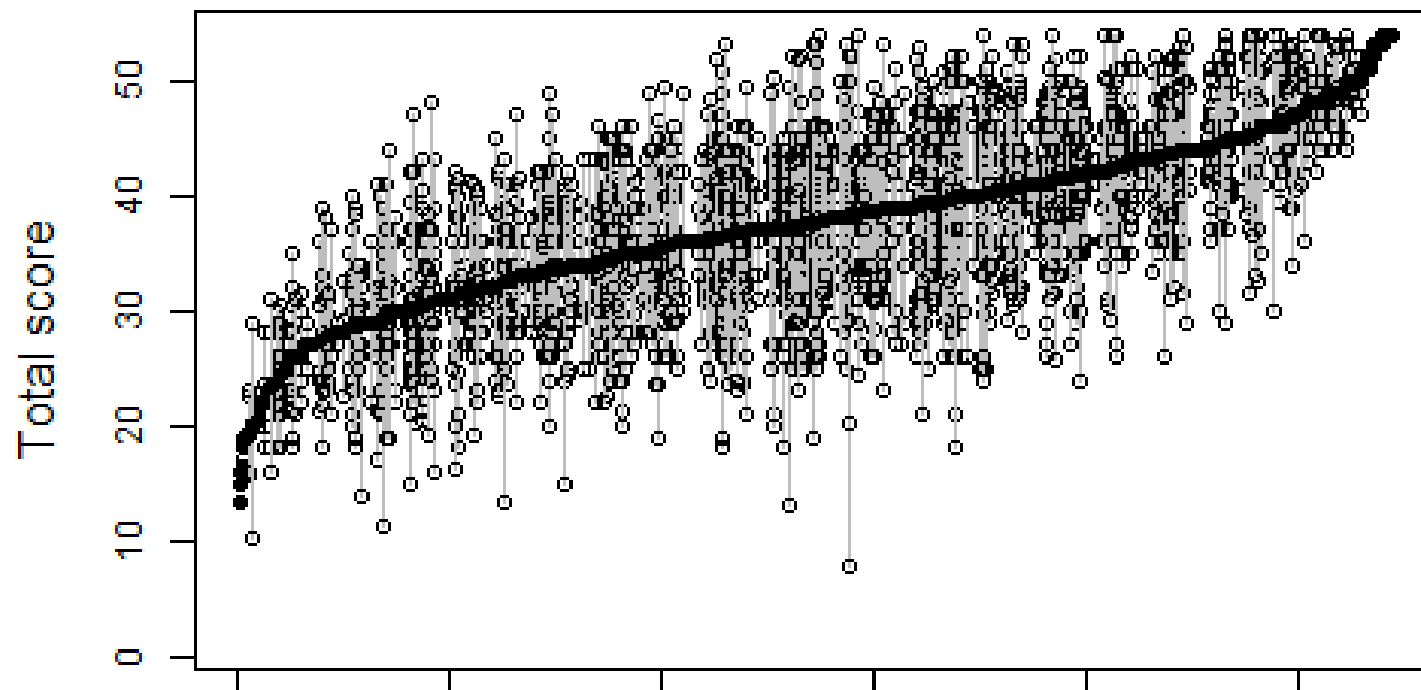


Applications ranked by average total score

Motivace: (Ne)konzistence hodnocení

16

Mean and range of ratings



Applicants ranked by average total score

Inter-rater reliability: Struktura dat

17

- 3986 hodnocení (vyplněných formulářů)
- 1 177 uchazečů
 - hodnoceni 1-25 krát
 - hodnoceni pro 1-17 škol
 - interní a externí
- 141 hodnotitelů
 - hodnotili 1-99 times
 - hodnotili uchazeče pro 1-8 škol
- 526 inzerátů na 54 školách
 - ZŠ, SŠ
 - 15 typů pozic (ZŠ: třídní, SŠ: matematika, angličtina, ...)

Inter-rater reliability: Cíle

18

- **Odhad Inter-Rater Reliability**
 - Globálně, na úrovni škol, na úrovni inzerátů
 - Porovnat IRR pro jednotlivé komponenty
- **Testovat, co může způsobovat nekonzistence**
 - Vhodnost uchazeče pro danou školu (resp. pro daný inzerát)
 - Je IRR menší pro externí uchazeče?
 - Je IRR menší pro SŠ než pro ZŠ?
- **Implikace**
 - Jak lze zvýšit reliabilitu pomocí průměrování více hodnotitelů
 - Jak lze zvýšit predikční validitu

Odhad Inter-Rater Reliability

19

Tradiční odhady:

- Založené na korelačním koeficientu nebo
- Založené na analýze rozptylu
- Vyžadují vyvážený a poměrně jednoduchý design

Náš přístup:

- Hierarchické lineární modely (modely se smíšenými efekty) pomocí knihovny `lme4` v statistickém prostředí `R`
- Odhady metodou reziduální max. věrohodnosti
- Parametrický bootstrap pro výpočet 95% konf. int. IRR
- Výběh optimálního modelu pomocí BIC

Odhad IRR: Napříč školami

20

Model:

$$Y_{ij} = \mu + A_i + B_j + e_{ij}$$

Efekt uchazeče $A_i \sim N(0, \sigma_A^2)$

Efekt hodnotitele $B_j \sim N(0, \sigma_B^2)$

Náhodná chyba $e_{ij} \sim N(0, \sigma_e^2)$

Reliabilita:

$$R = \frac{\textit{variance skutečných skóru}}{\textit{celková variance}} = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_B^2 + \sigma_e^2}$$

Odhad IRR: Uvnitř jednotlivých škol

21

Model na úrovni jednotlivých škol:

$$Y_{ijk} = \mu + A_i + B_j + AS_{ik} + e_{ijk}$$

Reliabilita:

$$R = \frac{\textit{variance skutečných skóre}}{\textit{celková variance}} = \frac{\sigma_A^2 + \sigma_{AS}^2}{\sigma_A^2 + \sigma_B^2 + \sigma_{AS}^2 + \sigma_e^2}$$

Odhad IRR: Uvnitř výběrových řízení

22

Model na úrovni jednotlivých výběrových řízení (inzerátů):

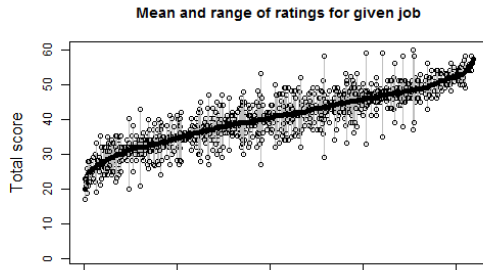
$$Y_{ijkl} = \mu + A_i + B_j + AS_{ik} + AJ_{il} + e_{ijkl}$$

Reliabilita:

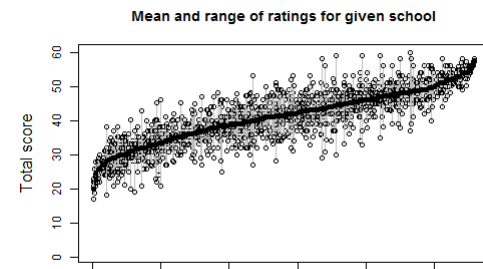
$$R = \frac{\textit{variance skutečných skóre}}{\textit{celková variance}} = \frac{\sigma_A^2 + \sigma_{AS}^2 + \sigma_{AJ}^2}{\sigma_A^2 + \sigma_B^2 + \sigma_{AS}^2 + \sigma_{AJ}^2 + \sigma_e^2}$$

Odhad IRR na třech úrovních

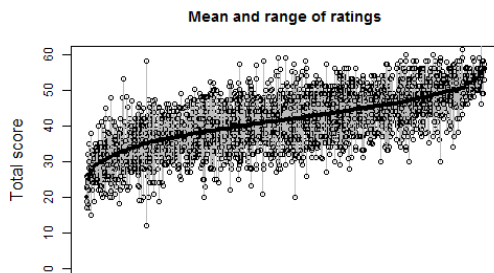
23



Applications ranked by average total score



Applications ranked by average total score

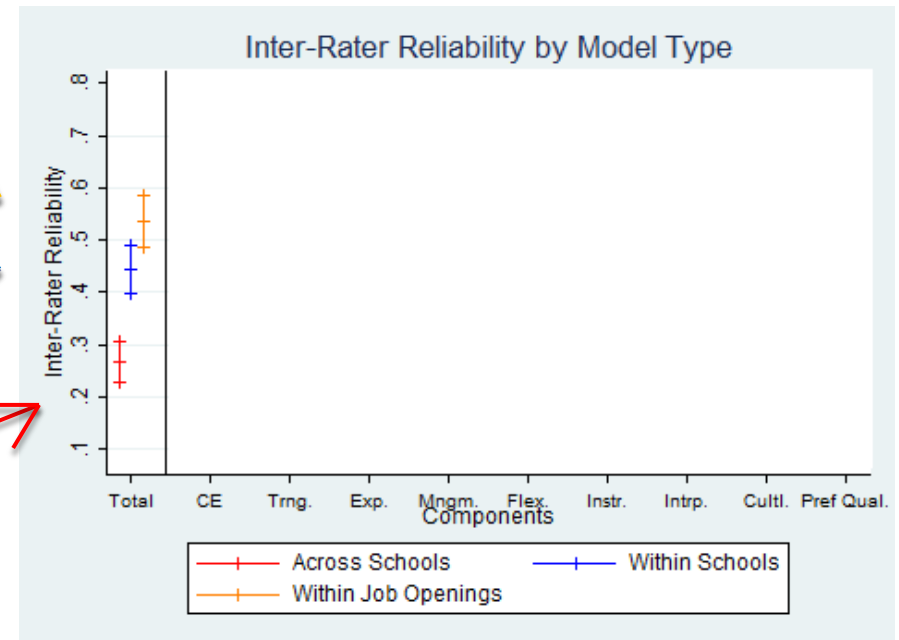


Applications ranked by average total score

3. Uvnitř
jednotlivých
výběrových řízení
(inzerátů)

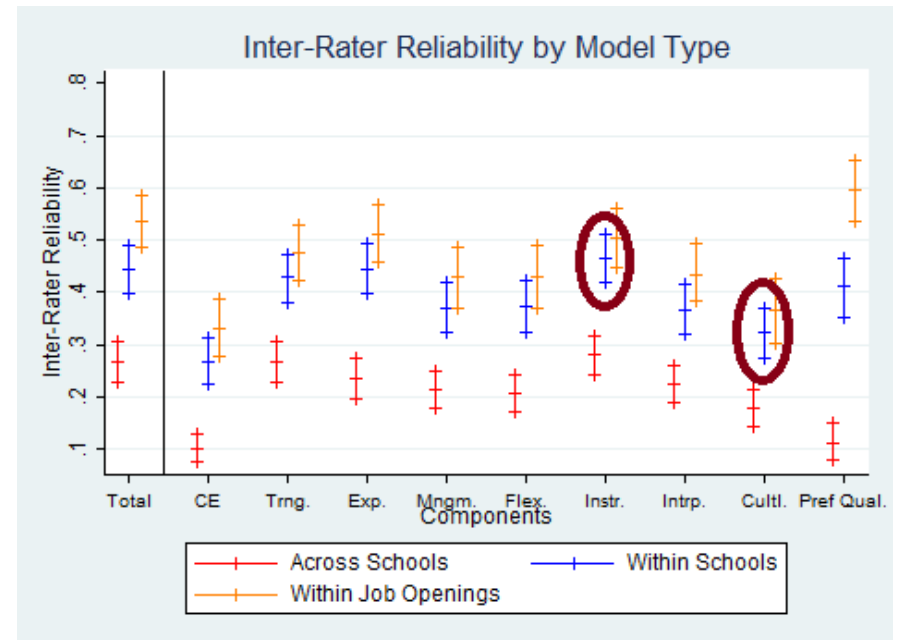
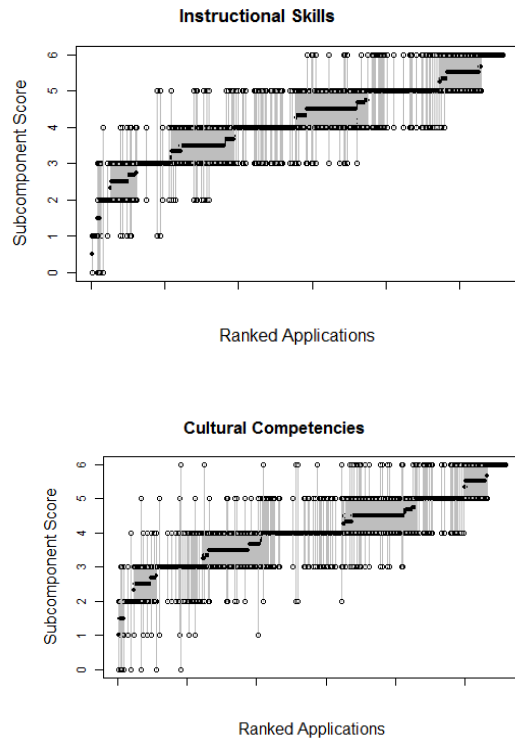
2. Uvnitř
jednotlivých škol

1. Globální
(napříč školami)



Odhad IRR: Komponenty

24

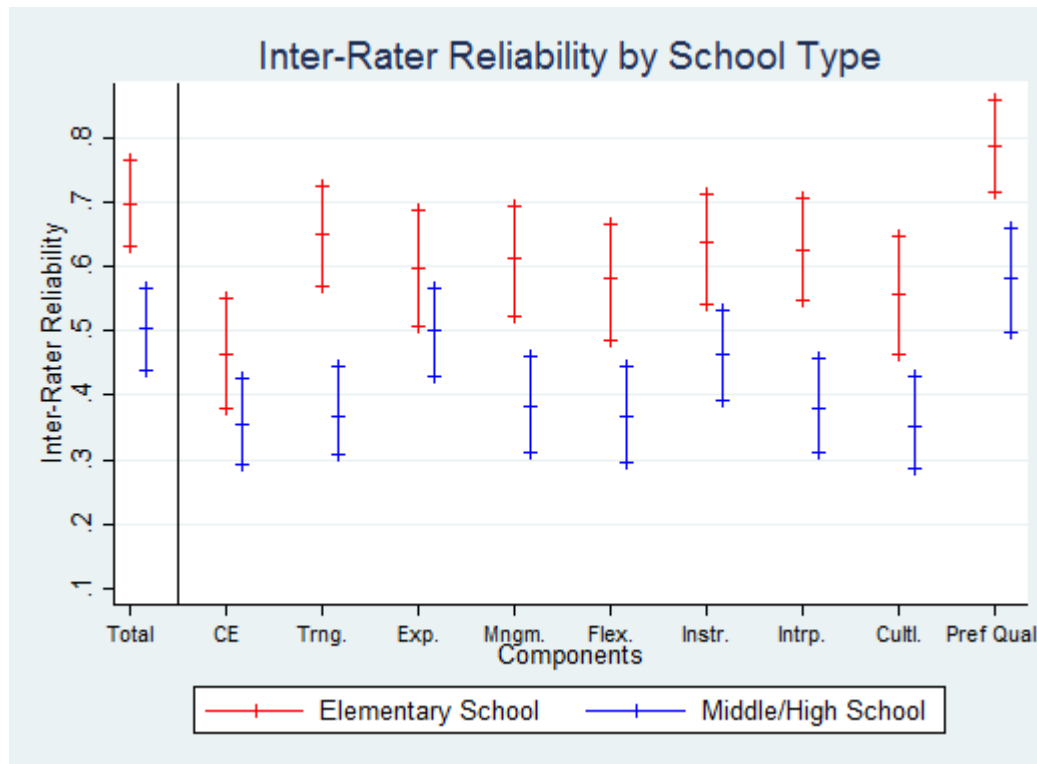


- Pro všechny komponenty jsou hodnocení závislé na škole, často i na výb. řízení (posuzováno dle BIC pro jednotlivé modely)
- Některé komponenty jsou míň spolehlivé než jiné (např. Kulturní kompetence vs. Instruktažní dovednosti)

Co způsobuje nízkou IRR?

25

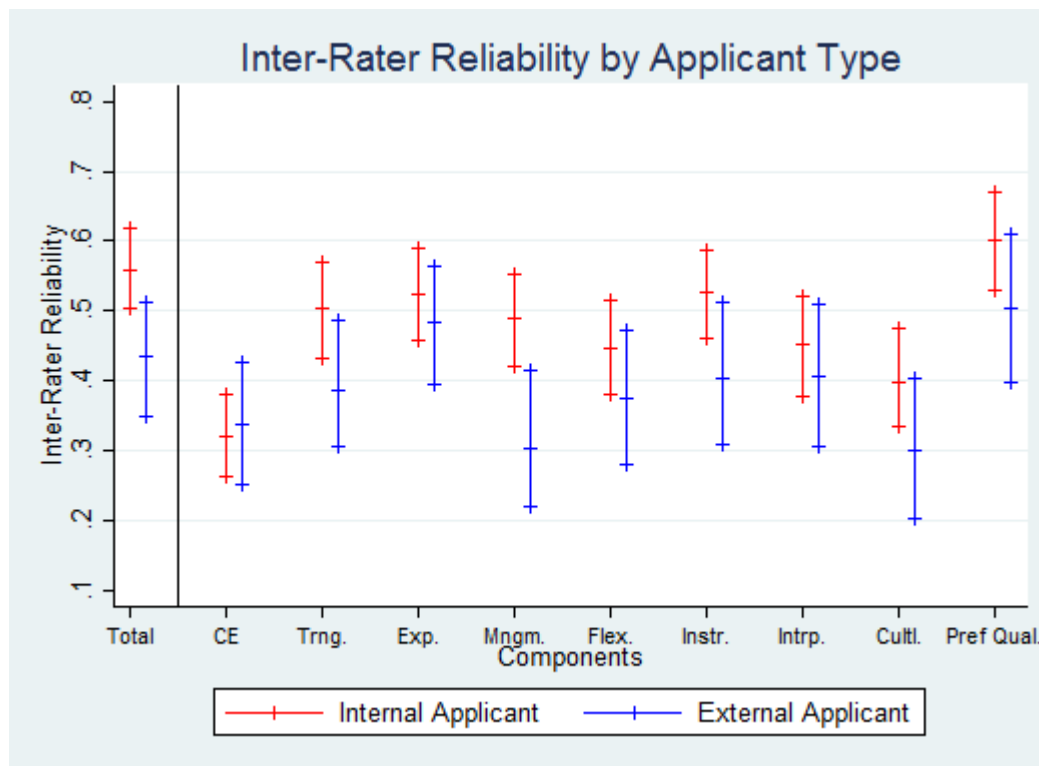
Typ školy: nižší Inter-Rater Reliabilita pro SŠ



Co způsobuje nízkou IRR?

26

Typ uchazeče: nižší Inter-Rater Reliabilita pro externí uchazeče



Možnosti zvýšení inter-rater reliability

27

- Průměrování dvou a více hodnotitelů
 - Dva hodnotitelé stačí k zvýšení IRR na 0,65 u některých komponent
 - Tři hodnotitelé stačí ke zvýšení IRR na 0,80
- Trénink a certifikace hodnotitelů
- Zveřejnění kritérií těm, kdo píší doporučující dopisy
- Důvěrnost doporučujících dopisů
- Přidání desetinného místa

Přímý vliv na predikční validitu

28

- Vysoká reliabilita je nutnou podmínkou pro validitu
 - Kulturní kompetence
- Vyšší reliabilita nezaručuje lepší predikční schopnost
 - Instruktažní schopnosti vs. Vedení třídy (Classroom management)
- Průměrování dvou hodnotitelů zvýší korelaci s přidanou hodnotou učitele (value added) asi o 20%

Závěry, diskuse

Závěry, diskuse

30

- Přijímací řízení je klíčovým momentem pro instituci i uchazeče
- Strukturované přijímací řízení s jasně stanovenými kritérii má potenciál predikovat úspěšnost uchazeče
- Predikční validita je úzce svázaná s konzistencí hodnotitelů (IRR)
- K odhadu reliability ve složitějších designech lze využít hierarchických (smíšených lineárních) modelů
- Další využití:
 - Ústní pohovory a další hodnocení v přijímacích zkouškách na VŠ
 - Observační hodnocení učitelů, studentů, škol
 - Hodnocení návrhů projektů GA ČR aj.

Děkuji za pozornost!

Dotazy, náměty?

martinkova@cs.cas.cz

www.cs.cas.cz/martinkova

Reference

- Goldhaber, Grout & Huntington-Klein: *Screen Twice, Cut Once: Assessing the Predictive Validity of Teacher Selection Tools*. CEDR WP #2014-9
- Martinková & Goldhaber: *Improving Teacher Selection: The Effect of Intra-Rater Reliability in the Screening Process*. CEDR WP #2015-7
- Martinková & Vlčková: *Hodnocení reliability znalostních a psychologických testů*. Informační bulletin České statistické společnosti, 25(4), 2014, str. 1-15.

BACKUP SLIDES

Data Overview

34

- Spokane Public Schools: applicant profile data, screening scores, record of job applications and status, and teacher absence data from 2008-09 to 2012-13 SYs (4 years of applicant data)
- Washington State OSPI
 - S-275 personnel records for certificated public school positions
 - PESB certification and licensure exam records
 - Core Student Record System: student achievement, demographic and assignment information
- NCES: school and district-level data
- WSIPC: teacher absence data for other districts (date and type)

Data Collected During Application Process

Table 1. Teacher Applicant Data

Applicant Background	Applicant Experience
Application Date	Student Teaching Experience
Full Name	Start Date
Employee ID	End Date
Address	State
Overall GPA	District
Race/Ethnicity	Subject
Degree Type	Title I Funded? (Y/N)
Degree Major	Teaching Experience
Degree College	Start Date
Degree Date	End Date
Degree Start Date	State
Degree End Date	District
Certification Area	Position Title
Certification Field	Subject
Certification Type	Reason Left
Certification State	Non-Teaching Experience
	Start Date
Pre-Screen Score (21 Pt Rubric)	End Date
Screening Score (54 Pt Rubric)	Company
	Title
	Hours
	City
	State
	Reason Left

Note: The applicant data accommodates multiple degrees, student teaching experiences, teaching experiences, and non-teaching experiences.