



Technický manuál

verze 2.0

Kevin S. McGrew ♦ Erica M. LaForte ♦ Fredrick A. Schrank

Tomáš Urbánek (Ed.)

Spoluautoři české verze

Anton Furman ♦ Lenka Krejčová ♦ Jiří Laciga ♦ Monika Víchová

Autoři českých dílčích klinických validizačních studií

Adam Klocek ♦ Martina Schimová ♦ Martina Vítková ♦ Hynek Cígler

Spolupráce na statistických analýzách

Hynek Cígler ♦ Edita Chvojková ♦ Petr Palíšek ♦ Adam Ťápal

Autoři původních textů v originálním technickém manuálu

Kevin S. McGrew ♦ Erica M. LaForte ♦ Fredrick A. Schrank

Reference Citations

- To cite the entire WJ IV battery, use:
Schrack, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock-Johnson IV*. Rolling Meadows, IL: Riverside.
- To cite the WJ IV Tests of Achievement, use:
Schrack, F. A., Mather, N., & McGrew, K. S. (2014). *Woodcock-Johnson IV Tests of Achievement*. Rolling Meadows, IL: Riverside.
- To cite the WJ IV Tests of Cognitive Abilities, use:
Schrack, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock-Johnson IV Tests of Cognitive Abilities*. Rolling Meadows, IL: Riverside.
- To cite the WJ IV Tests of Oral Language, use:
Schrack, F. A., Mather, N., & McGrew, K. S. (2014). *Woodcock-Johnson IV Tests of Oral Language*. Rolling Meadows, IL: Riverside.
- To cite this manual, use:
McGrew, K. S., LaForte, E. M., & Schrack, F. A. (2014). Technical Manual. *Woodcock-Johnson IV*. Rolling Meadows, IL: Riverside.
- To cite the online scoring and reporting program, use:
Schrack, F. A., & Dailey, D. (2014). *Woodcock-Johnson Online Scoring and Reporting* [Online format]. Rolling Meadows, IL: Riverside.

Copyright © 2014 by The Riverside Publishing Company. No part of this work may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording or by any information storage or retrieval system, without the prior written permission of Riverside, unless such copying is expressly permitted by federal copyright law. Address all inquiries to Permissions, Riverside, 3800 Golf Road, Suite 200, Rolling Meadows, Illinois 60008.

Bateria III Woodcock-Muñoz, WJ III, WJ-R, and Woodcock-Johnson are registered trademarks of Houghton Mifflin Harcourt Publishing Company. WJ IV and the Woodcock-Johnson logo are trademarks of Houghton Mifflin Harcourt Publishing Company.

CELF, Clinical Evaluation of Language Fundamentals, DAS-II, Differential Ability Scales, Wechsler Adult Intelligence Scale, WAIS, Wechsler Individual Achievement Test, WIAT, Wechsler Intelligence Scale for Children, and WISC are registered trademarks of Pearson Education, Inc.

CASL, KTEA, OWLS, PPVT, Wechsler Preschool and Primary Scale of Intelligence, and WPPSI are trademarks of Pearson Education, Inc.

Ping-Pong is a registered trademark of Escalade Sports.

The MindHub is a registered trademark of the Institute for Applied Psychometrics (IAP) and Interactive Metronome.

WINSTEPS is a registered trademark of John Michael Linacre.

All other trademarks are the property of their respective owners.

The WJ IV tests are not to be used in any program operating under statutes or regulations that require disclosure of specific item content and/or correct responses to the public, including subjects or their parents. Any unauthorized distribution of the specific item content and/or correct responses is prohibited by copyright law.

For technical information, please visit www.riversidepublishing.com or call Riverside Customer Service at 800.323.9540.

First published by The Riverside Publishing Company,
3800 Golf Road, Suite 200, Rolling Meadows, Illinois, 60008-4015, United States of America

Copyright English version © 2014 by The Riverside Publishing Company.

Translation © 2019 by Riverside Assessments, LLC.

One Pierce Place, Suite 900W, Itasca, Illinois, 60143, United States of America. All rights reserved

Žádná část tohoto díla nesmí být rozmnožována, ukládána ani přenášena elektronickými, mechanickými, kopírovacími, filmovacími, záznamovými či jinými prostředky bez písemného povolení vydavatele.

Vydalo Propsyco s.r.o, Bartošova 316, Otrokovice

Tomáš Urbánek (Ed.)

Spoluautoři české verze: Anton Furman, Lenka Krejčová, Jiří Laciga, Monika Víchová

Autoři českých dílčích klinických validizačních studií: Adam Klocek, Martina Schimová, Martina Vitková, Hynek Cígler

Překlad a korektury: Monika Víchová, Edita Chvojková, Petr Palíšek

Grafická úprava: Martin Šretr

Text je volně dostupný na www.propsyco.cz

O autorech WJ IV



Fredrick A. Schrank

Fredrick A. (Fred) Schrank působil na pozici vedoucího autorského týmu v čele vývoje Woodcock-Johnson® IV (WJ IV™). Řídil společnost Measurement Learning Consultants (MLC), která se zabývala vývojem, a poskytoval služby nadaci Woodcock-Muñoz Foundation.

Dr. Schrank je držitelem psychologické licence (Washington) a certifikovaným specialistou v oblasti školní psychologie od American Board of Professional Psychology (ABPP). Než získal titul Ph.D. na University of Wisconsin-Madison, působil ve školních okresech Dodgeville, North Fond du Lac a De Forest ve Wisconsinu. Poté vyučoval na Truman State University v Missouri a na University of Puget Sound ve Washingtonu, následně pak věnoval 25 let života téměř výhradně vývoji a publikování testů systému Woodcock-Johnson. Jako profesionální psycholog byl examínátorem pro American Board of School Psychology (ABSP) a prezidentem American Academy of School Psychology (AASP). Fred Schrank hrál rovněž ústřední roli při tvorbě plánu organizace a interpretace WJ IV, včetně online programu Woodcock-Johnson pro skórování a interpretaci.



Kevin S. McGrew

Kevin S. McGrew je ředitelem Institute for Applied Psychometrics (IAP), LLC, což je nezávislá výzkumná a poradenská organizace založená v r. 1998. Byl rovněž pomocným ředitelem společnosti Measurement Learning Consultants a vedoucím vývoje v nadaci Woodcock-Muñoz Foundation. Hostuje též jako přednášející pedagogické psychologie v programu školní psychologie na University of Minnesota. Je vedoucím výzkumu ve společnosti Interactive Metronome, která se zabývá neurotechnologiemi a rehabilitací. Získal titul Ph.D. v oblasti pedagogické psychologie (speciální pedagogika) na University of Minnesota, titul MS ve školní psychologii a BA v psychologii na Minnesota State University-Moorhead.

Dr. McGrew působil 12 let jako školní psycholog ve státech Iowa a Minnesota. V období 1989 až 2000 byl profesorem na Katedře aplikované psychologie na St. Cloud State University v St. Cloud ve státě Minnesota. Byl poradcem v oblasti měření pro řadu vydavatelů psychologických testů, celostátních výzkumných studií a organizací.

Dr. McGrew je autorem množství publikací, příspěvků na státních, celonárodních a mezinárodních konferencích na témata související s jeho výzkumnými zájmy o lidskou inteligenci, hodnocení intelektu, lidské kompetence, aplikovanou psychometrii a Cattell-Horn-Carrollovu (CHC) teorii kognitivních schopností. Aktivně sdílí teoretické a výzkumné poznatky prostřednictvím tří profesionálních blogů a internetového portálu The MindHub®.

Dr. McGrew byl hlavním poradcem v otázkách měření při tvorbě WJ-R a ve stejné pozici působil i jako spoluautor *Mini-Battery of Achievement* (MBA), *Sharpe-McNear-McGrew Braille Assessment Inventory* (BAI), WJ III, *Woodcock-Johnson Diagnostic Supplement to the Tests of Cognitive Abilities*, *Batería III Woodcock-Muñoz®* (Batería III), *Woodcock-Johnson III Normative Update*, *Woodcock-Johnson III* – australská verze a WJ IV. Působil rovněž jako poradce pro otázky psychometrie a statistiky při vývoji *Children's Psychological Processes Scale*.



Nancy Matherová

Nancy Matherová působí jako profesorka na Katedře poruch učení a jejich edukace na University of Arizona. Získala titul MA v oblasti poruch chování a Ph.D. v oboru speciálního vzdělávání a poruch učení na University of Arizona. Na stejné univerzitě pak absolvovala postdoktorální stipendijní pobyt pod vedením Dr. Samuela Kirka.

Dr. Matherová pomáhala Dr. Richardovi Woodcockovi s několika aspekty vývoje testu pro *Woodcock-Johnson Psycho-Educational Battery-Revised (WJ-R®)*, mezi něž patřilo i spoluautorství manuálu administrátora pro testy kognitivních schopností a úspěšnosti WJ-R. Autorsky se rovněž podílela na *Woodcock-Johnson III (WJ III®)* a WJ IV, je také spoluautorkou dvou knih zaměřených na interpretaci a aplikaci WJ III – *Essentials of WJ III Tests of Achievement Assessment a Woodcock-Johnson III: Reports, Recommendations, and Strategies*.

Působila jako vyučující poruch učení, diagnostička, univerzitní profesorka a poradkyně pro vzdělávání. Dr. Matherová provádí výzkum v oblasti rozvoje čtení a psaní. Publikovala množství článků, vede workshopy o diagnostice a výuce na celostátní i mezinárodní úrovni, je spoluautorkou několika publikací propojujících oblast hodnocení a intervence, mezi něž patří *Learning Disabilities and Challenging Behaviors: A Guide to Intervention and Classroom Management*, *Evidence-Based Interventions for Students with Learning and Behavioral Challenges*, *Essentials of Assessment Report Writing*, *Essentials of Evidence-Based Academic Interventions*, *Writing Assessment and Instruction for Students with Learning Disabilities* a nejnovější *Essentials of Dyslexia: Assessment and Intervention*.

Spolupracující autor

Erica M. LaForte je spoluautorkou Technického manuálu WJ IV s Kevinem McGrewem a Fredrickem Schrankem. Je odborníci v oboru testování v oblasti vzdělávání s více než 15 lety zkušeností s vývojem a publikováním testů v oblasti klinické a školní psychologie a testů pro udělování licenci a certifikátů. Získala titul MA z Výzkumné metodologie na Loyola University-Chicago a titul PhD z psychologie ve vzdělávání (se zaměřením na měření, hodnocení, statistiku a diagnostiku) na University of Illinois v Chicagu. V průběhu vývoje WJ IV poskytovala Erica základní vedení psychometrických analýz položek, škálování, vyrovnávání testů a analýz reliability testů a klastrů. Asistovala autorskému týmu také při analýzách souběžné validity a klinických studiích. Než začala pracovat na WJ IV, byla Erica 13 let zaměstnána v nakladatelství Riverside, kde zastávala různé pozice související s vývojem a výzkumem testů.

Použité nástroje

Při zpracování dat byl použit programovací jazyk R a (použité knihovny a jejich verze jsou uvedeny v seznamu literatury) a program Winsteps (Linacre, 2012).

O autorech české adaptace WJ IV

PhDr. A. Furman, CSc. začal svoji profesionální dráhu v Krajské pedagogicko-psychologické poradně v Banské Bystrici. Později pracoval ve Výzkumném ústavu dětské psychologie a patopsychologie a krátce i v Slovenské akademii věd v Bratislavě, odkud se vrátil do Banské Bystrice vyučovat psychologii na Pedagogické fakultě Univerzity Mateja Bela. Byl zakládajícím členem a prezidentem Asociace školní psychologie SR a ČR a poté i prezidentem Mezinárodní asociace školní psychologie. Po emigraci do USA v roce 1998 začal aktivně spolupracovat s Richardem Woodcockem a jeho týmem na adaptaci kognitivní části baterie testů Woodcock-Johnson® do vícero jazyků, včetně českého a slovenského. Práci školního psychologa ve státě Washington kombinuje s trenérskou aktivitou zaměřenou na uživatele testů Woodcock-Johnson.

PhDr. Lenka Krejčová, Ph.D. působí jako psycholožka v neziskové organizaci DYS-centrum Praha z. ú. a jako odborná asistentka na Katedře psychologie FF UK, kde přednáší pedagogickou psychologii. V DYS-centru se věnuje klientům od předškolního věku po dospělost, kteří čelí různorodým výukovým obtížím. Poskytuje zde konzultace, dlouhodobé terapie, komplexní diagnostiku. Mezi její profesní zájmy patří specifické poruchy učení, kognitivní psychologie, zejména s důrazem na proces učení, diagnostika, dynamická diagnostika. Podílela se na lokalizaci několika diagnostických nástrojů (např. ACFS, CAS2), je zapojena do mezinárodních projektů i výzkumných studií, především v souvislosti s procesy učení. Je autorkou či spoluautorkou řady monografií i odborných článků.

PhDr. Jiří Laciga, Ph.D. provozuje privátní klinickou praxi. Poskytuje zde konzultace, dlouhodobé terapie, komplexní diagnostiku. Věnuje se tvorbě a adaptaci intelektových a paměťových škál, je v této oblasti zapojen do výzkumných studií. Je hlavním autorem české adaptace Testu paměti a učení (TOMAL 2). Publikoval v domácích i zahraničních odborných časopisech.

Prof. PhDr. Tomáš Urbánek, Ph.D. je psycholog zabývající se výzkumem, především v oblasti metodologie, psychometrických analýz v základním a aplikovaném výzkumu a psychosémantiky. Je autorem a spoluautorem několika desítek vědeckých článků, několika knih, a spoluautorem několika psychodiagnostických metod (např. IDS, CAST, CFT 20-R). Působí v Psychologickém ústavu AV ČR a Psychologickém ústavu FF MU v Brně.

Mgr. et Mgr. Monika Víchová je absolventkou Filozofické fakulty Masarykovy univerzity v oboru Psychologie a v oboru Český jazyk a literatura. Aktuálně v Psychologickém ústavu FF MU dokončuje Ph.D. studium v oboru Obecná psychologie. Výzkumně se věnuje psycholinguistice, psychodiagnostice a problematice duševních onemocnění. Pracuje jako psycholog ve zdravotnictví v psychiatrické nemocnici.

Poděkování

Česká verze WJ IV COG vznikla za přispění desítek osob.

Na přípravě nahrávek, školení examinátorů a korekturách testu se významně podílela PhDr. Jarmila Burešová.

K psychometrickému zpracování významně svými podněty přispěla dr. Erica LaForte.

Další cenné připomínky ke statistickým analýzám poskytl Mgr. Hynek Cígler, Ph.D., který také připravil řadu skriptů pro vyhlazování norem.

Sběr dat organizovala Mgr. Adéla Rajsiglová.

Kompletní kódování provedených standardizačních vyšetření pořídila Magdalena Pišanová.

Online vyhodnocovací program vytvořil Ing. Miroslav Zárecký.

Za grafické zpracování vděčíme Martinu Šretrovi.

Předmluva

Tento technický manuál má sloužit psychometricky vzdělanému čtenáři jako další příručka, ve které najde podrobnosti týkající se české adaptace testu Woodcock-Johnson® IV Testy kognitivních schopností. Jeho struktura vychází do značné míry z americké verze, která je volně k dispozici na webu (McGrew, LaForte, Schrank, 2014). Je však podstatně stručnější, protože zatímco americká verze testu se skládá ze tří baterií – Testů kognitivních schopností, Testů mluveného jazyka a Testů školních dovedností, česká verze obsahuje zatím pouze Testy kognitivních schopností.

V průběhu standardizace bylo nutné provést stovky dílčích analýz dat z jednotlivých testů, jejichž výstupy by zabraly mnoho set stran a které by měly cenu asi jen pro úzce psychometricky orientované čtenáře. Řada výsledků by ale kvalifikovaného čtenáře zajímat měla – např. detaily týkající se popisných statistik a podmíněných standardních chyb měření, ze kterých jsou odvozeny intervaly spolehlivosti skóre, vzájemné korelace testů a jejich faktorová struktura atd. Tento technický manuál tyto výstupy uvádí.

Původní americký technický manuál obsahuje texty, které jsou relevantní i pro uživatele české verze testové baterie. Jde jednak o obecné texty týkající se CHC teorie inteligence, Raschových škál, které byly použity při ověřování psychometrických vlastností jednotlivých testů, a jednak o pasáže věnované vývoji americké verze testové baterie, které mohou být užitečné i pro české uživatele. Tyto texty byly přeloženy do češtiny a použity v tomto manuálu – obecnější texty v plném rozsahu, specifické texty byly zpravidla kráceny a doplněny informacemi týkajícími se české adaptace testové baterie. Autorství jednotlivých částí je vyznačeno a stejně tak jsou uvedeni autoři překladů, kterými byli Monika Víchová, Edita Chvojková a Petr Palíšek.

Obsah

O autorech WJ IV	iii
Použité nástroje	v
O autorech české adaptace WJ IV	v
Poděkování	vi
Předmluva	vii
Kapitola 1 Přehled WJ IV	1
Cíle a vývoj revidované verze WJ IV	1
Vývoj CHC teorie ve WJ IV	2
Revidované nebo nové CHC široké a úzké schopnosti ve WJ IV	3
Krátkodobá pracovní paměť (Gwm)	3
Paměť na zvukové vzorce (UM)	4
Kognitivní komplexita	4
Auditivní zpracování (Ga)	5
Organizace do tří odlišných baterií	6
WJ IV Testy kognitivních schopností	6
Návrhy použití a interpretace testových skóre WJ IV	7
Kapitola 2 Design testu a postup vývoje	8
Aktualizace testových specifikací	8
Vývoj nových testů WJ IV	8
Vývoj české verze	10
Test 1: Slovník	10
Test 3: Verbální pozornost	11
Test 5A: Fonologické zpracování – Vyhledávání slov	11
Test 5B: Fonologické zpracování – Slovní fluence	11
Test 5C: Fonologické zpracování – Substituce	12
Test 6: Reprodukce příběhů	12
Test 8: Všeobecné znalosti	12
Test 12: Opakování pseudoslov	13
Test 13: Audio-vizuální učení	13
Test 16: Řazení názvů a čísel	14
Test 18: Paměť na slova	14
Vývoj škály	14
Využití Raschova modelu	15
Škála W	16
Vývoj položkových bank	17
Ověřovací studie	18

Kalibrace položkových bank	18
Sestavování testů pro tvorbu norem	18
Sběr dat pro tvorbu norem	19
Kalibrace a vyvažování dat z normativní studie	19
Kalibrace časovaných testů	20
Analýzy zkreslení položek	21

Kapitola 3 Standardizace a vývoj norem WJ IV **23**

Obecné charakteristiky normalizačního souboru	23
Postupy normalizační studie	24
Tvorba norem	24
Souhrnné statistiky testů a klastrů	25
Tvorba a výpočet klastrových skóre WJ IV	25
Výpočet diferenciálně váženého klastrového skóre GIA	25
Konstrukce norem a odvozených skóre WJ IV	25
Popis tradičního postupu výpočtu norem	26
Výhody novějších metod výpočtu hodnot REF W	30
Výpočet věkových ekvivalentních skóre	31
Výpočet percentilových pořadí a standardních skóre	32
Výpočet rozdílových skóre WJ IV	32

Kapitola 4 Reliabilita **35**

Koncepty reliability	35
Chyba měření	36
Koeficienty reliability	37
Reliability testů	38
Testy a subtesty s dichotomicky skórovanými položkami	38
Testy s vícebodovým skórováním	38
Testy se subtesty	39
Testy s rychlostní složkou	40
Reliability klastrů	41

Kapitola 5 Důkazy o validitě WJ IV **42**

Reprezentativnost obsahu baterie WJ IV a její procesuální a konstruktové pokrytí	42
Pokrytí teorie CHC	43
Pokrytí konstruktů, procesů a obsahu	44
Empirické zhodnocení charakteristik obsahu testů	47
Kognitivní komplexita obsahu a operací	49
Vývojové křivky klastrů schopností WJ IV	52
Obecné (g), široké a úzké shluky WJ IV Cog	54
Vývojové křivky v datech české standardizace	54
Vnitřní struktura WJ IV	56
Interkorelace testů a klastrů ve standardizačním výběru	56
Důkazy o vnitřní struktuře: Trojstupňová analýza strukturní validity WJ IV	57
Fáze 1: Náhodné rozdělení vzorku	58
Analýza strukturní validity ve všech šesti věkových kohortách	58
Fáze 2: Explorační generování strukturních modelů (MG) a jejich ověření na souborech pro vývoj modelu	59
Fáze 2A: Explorační shluková analýza	59
Fáze 2A: Explorační analýza hlavních komponent (PCA)	61

Fáze 2A: Explorační multidimenzionální škálování	62
Fáze 2B: Specifikace počátečních modelů CFA	63
Fáze 3: Cross-validizace konfirmačních strukturních modelů	63
Zhodnocení fitu	63
Důkazy o vnitřní struktuře: výsledky a interpretace důkazů o strukturní validitě	64
Srovnání shody modelů CFA s daty	64
Vztah skóre WJ IV s jinými nástroji na měření kognitivních schopností	64
Výkon ve WJ IV u osob z klinických vzorků	65
Doklady strukturní validity založené na analýzách dat české standardizace	65
Hierarchická shluková analýza	73
Multidimenzionální škálování	75
Validizační studie pro vybrané klinické skupiny	78
Výběrové soubory:	78
Proces administrace a úprava administrace u klinických skupin:	79
Obecná doporučení k administraci u klinických skupin:	79
Specifická doporučení pro testování člověka s úzkostí, depresí a schizofrenií:	80
Stručné výsledky jednotlivých diplomových prací	81

Literatura 83

Příloha A Definice Cattellovy-Hornovy-Carrollovy (CHC) teorie kognitivních schopností 92

Doménově nezávislé obecné kapacity	92
Fluidní inteligence (Fluid reasoning, <i>Gf</i>)	92
Krátkodobá pracovní paměť (Short-Term Working Memory; <i>Gwm</i>)	93
Dlouhodobá paměť a vybavení (Long-Term Storage and Retrieval; <i>Glr</i>)	93
Glr-Efektivita učení	93
Glr-Fluence vybavení	94
Rychlost zpracování (Processing speed; <i>Gs</i>)	94
Rychlost reakcí a rozhodování (Reaction and Decision Speed; <i>Gt</i>)	95
Psychomotorické tempo (Psychomotor speed, <i>Gps</i>)	95
Systémy získaných znalostí	96
Porozumění-Znalosti (Comprehension-Knowledge; <i>Gc</i>)	96
Doménově specifické znalosti (Domain-Specific Knowledge; <i>Gkn</i>)	96
Čtení a psaní (Reading and Writing; <i>Grw</i>)	97
Kvantitativní znalosti (Quantitative knowledge; <i>Gq</i>)	98
Senzorické/Motoricky podmíněné schopnosti (Sensory/Motor-Linked Abilities)	98
Vizuální zpracování (Visual processing, <i>Gv</i>)	98
Auditivní zpracování (Auditory processing; <i>Ga</i>)	99
Olfaktorické schopnosti (Olfactory abilities, <i>Go</i>)	100
Taktilní schopnosti (Tactile abilities, <i>Gh</i>)	100
Kinestetické schopnosti (Kinesthetic abilities; <i>Gk</i>)	100
Psychomotorické schopnosti (Psychomotor abilities, <i>Gp</i>)	101

Příloha B Popisné statistiky jednotlivých testů 102

Příloha C Popisné statistiky klastrových skóre 111

Tabulky

Tabulka 1-1. WJ IV COG Tabulka selektivního testování ukazující testy a interpretační klastry	7
Tabulka 2-1. Pravděpodobnost úspěchu pro několik klíčových hodnot $W_A - D$	17
Tabulka 2-2. Procenta položek s potenciálním DIF	22
Tabulka 3-1. Věkové charakteristiky standardizačního souboru	24
Tabulka 3-2. Vzdělání rodičů a vzdělání respondentů	24
Tabulka 3-3. Váhy pro výpočet skóru GIA v české standardizaci	25
Tabulka 4-1. Vybrané ukazatele reliability z Testu 5 WJ IV COG: Fonologické zpracování	38
Tabulka 4-2. Intervaly spolehlivosti v rozpětí ± 1 SEM pro vybrané hodnoty reliability a velikosti výběru	38
Tabulka 4-3. Popisné statistiky a koeficienty testové-retestové reliability z americké studie rychlostních testů WJ IV	41
Tabulka 5-1. Porovnání autorských a nezávislých klasifikací testů WJ III ponechaných ve WJ IV a autorské klasifikace nových testů WJ IV	45
Tabulka 5-2. Obsah testů WJ IV COG a přehled procesů a konstruktů	45
Tabulka 5-3. Náboje WJ IV COG ve faktoru g na prvním nerotovaném faktoru získaném metodou hlavních os	51
Tabulka 5-4. Zkratky označující jednotlivé testy	66
Tabulka 5-5. Věkové segmenty používané při analýze	67
Tabulka 5-6. Náboje jednotlivých testů pro PCA s 1-3 komponentami	73

Přehled WJ IV¹

Test Woodcock-Johnson® IV ² (WJ IV®) (Schrunk, McGrew, & Mather, 2014a) sestává ze tří odlišných, konormovaných baterií: Testů kognitivních schopností (*Woodcock-Johnson IV Tests of Cognitive Abilities; WJ IV COG*) (Schrunk, McGrew, & Mather, 2014b), Testů mluveného jazyka (*Woodcock-Johnson IV Tests of Oral Language; WJ IV OL*) (Schrunk, Mather, & McGrew, 2014b) a Testů školních dovedností (*Woodcock-Johnson IV Tests of Achievement; WJ IV ACH*) (Schrunk, Mather, & McGrew, 2014a). Společně tyto tři nástroje tvoří komplexní systém pro měření obecné inteligence (g), jednotlivých kognitivních schopností, mluveného jazyka a školních dovedností napříč širokým věkovým spektrem. Normativní data americké verze byla vytvořena na rozsáhlém, reprezentativním vzorku 7 416 osob ve věku od 2 do 90+ let.

Tato kapitola předkládá popis hlavních rysů a diskuzi k revidované verzi WJ IV, její cíle a vývoj testu. Prostor je věnován také popisu postupu pro porovnávání skóre testů a klastrů a jejich relevanci k současné diagnostické praxi. Následující kapitoly poskytují informace ohledně psychometrických charakteristik normativních dat WJ IV a rozličných skóre, které test poskytuje.

Cíle a vývoj revidované verze WJ IV

WJ IV je teoreticky, strukturně a interpretačně revidovanou verzí *Woodcock-Johnson III* (WJ III) (Woodcock, McGrew, & Mather, 2001, 2007), která vznikla jako nástroj pro měření obecné inteligence; široké a úzké kognitivní schopnosti jsou definovány dle moderní Cattellovy-Hornovy-Carrollovy (CHC) teorie, včetně mluveného jazyka, čtení, matematických dovedností a psaného jazyka, školních doménově-specifických vloh a školních znalostí. Níže popsané cíle a vývoj revidované verze WJ IV jsou v souladu s tradičními zásadami testů Woodcock-Johnson, které se zaměřují na kvalitu a vylepšování CHC teorie, a to už od její první formulace ve *Woodcock-Johnson Psycho-Educational Battery* (WJ) a WJ III tak, že nabízí více možností administrace a interpretace, aby vyhovely současným psychodiagnostickým potřebám.

1. WJ IV je navržen tak, aby poskytl co nejmodernější model měření v souladu s vývojem CHC teorie lidských kognitivních dovedností, a to:
 - a. vytvářením nových testů a interpretačních klastrů³, založených na uskutečněných studiích i potřebách odborníků z praxe tak, aby bylo možné měřit ty nejdůležitější kognitivní, jazykové a školní schopnosti;
 - b. zaměřením na ekologickou a prediktivní validitu klíčových interpretačních klastrů v baterii kognitivních, jazykových i školních dovedností díky zvýšení kognitivní komplexity vybraných testů; a
 - c. nabídnutím nového fluidně-krystalizovaného (*Gf-Gc*) kombinovaného kognitivního skóru ve WJ IV COG pro porovnání s dalšími mírami kognitivních procesů, jazykových kompetencí a školních dovedností, aby bylo možné určit silné a slabé stránky v jednotlivých doménách.
2. WJ IV je rozdělen do tří odlišných konormovaných baterií, které lze používat nezávisle na sobě, či různě kombinovat, což přináší profesionálním examinátorům v praxi větší flexibilitu díky:

1 Autorkou překladu této části je Monika Vichová

2 Tento text popisuje americkou verzi. Z textu jsou vynechány pasáže týkající se jiných částí WJ IV než kognitivní baterie. Vzhledem ke gramatickým a stylistickým obtížím, které přinášejí dvě příjmení v názvu testu (navíc Woodcock je muž, Johnson je žena), snažíme se maximálně používat formu „Test Woodcock-Johnson“.

3 Jde o součtové skóre kombinující skóre vybraných testů WJ IV. Protože nejde o ustálený pojem, rozhodli jsme se pro použití počestěného tvaru „klastř“, abychom ho maximálně odlišili např. od shluků získaných statistickou analýzou dat.

- a. pochopení důležitosti jazykových schopností jako základního korelátu kognitivního a školního fungování, a tedy dostupnosti WJ IV OL examinátorům, kteří provádějí diagnostiku kognitivních schopností, školních výkonů nebo úrovně jazykových znalostí;
 - b. začlenění celkového indexu jazykových⁴ schopností do různých verzí WJ IV, včetně praktických doporučení, jak testy administrovat;
 - c. nabídce tří paralelních forem standardní baterie WJ IV ACH pro zamezení nadměrného vystavení testovaných osob položkám jakékoliv formy.
3. WJ IV nabízí nový a užitečný způsob porovnávání dovedností v rámci jedné baterie i napříč různými bateriemi:
 - a. nabídkou možnosti sledování silných a slabých stránek jedince v oblasti kognitivních, jazykových i školních schopností;
 - b. uspořádáním každé baterie pro snadné použití stanovením klíčových testů pro každou baterii, které mohou být použity jako prediktor pro výpočet souvisejících silných a slabých stránek mezi administrovanými testy a klastry; a
 - c. vytvořením nových klastrů doménově specifických školních schopností, které dovolují vytvořit účinný a validní odhad výkonu.
 4. WJ IV si zachovává svůj důraz na psychometrické kvality, jež jsou spojeny s předešlými bateriemi testů Woodcock-Johnson, a to:
 - a. poskytnutím nového, rozsáhlého, reprezentativního normativního vzorku vybraného z populace USA;
 - b. aktualizací položek a zjednodušením postupů administrace i interpretace;
 - c. posílením škálování u rychlostních testů; a
 - d. použitím nejmodernějších způsobů sběru dat, vývoje testů a metod analýz dat jako modelových snadnějších postupů v oblasti aplikovaného vývoje testů.

Vývoj CHC teorie ve WJ IV

WJ IV představuje čtvrtou generaci komplexní baterie psychologicko-vzdělávacích testů, které začaly v roce 1973 a byly poprvé publikovány jako *Woodcock-Johnson Psycho-Educational Battery* (WJ) (Woodcock & Johnson, 1977). Zpočátku nebyly podloženy žádným teoretickým modelem, na jehož základě by byly vyvíjeny. WJ byl založen na pragmatickém rozhodovacím modelu (*decision-making model*) (McGrew, 1986; 1994; Woodcock, 1984). Formativní výzkum sestával z množství kontrolovaných experimentů zabývajících se měřením schopností učení. Sestavení testu vycházelo z vědecko-empirických metod. WJ obsahoval nesourodou směs úloh navrženou tak, aby pokryla kontinuum od nižších mentálních procesů (jednoduché operace) po vyšší mentální procesy (komplexní operace). Faktorové a shlukové analýzy byly prováděny z toho důvodu, aby pomohly určit menší počet širších schopností měřených baterií (Woodcock, 1978). Prvotní analýzy by mohly být označeny za okamžik, kdy bylo Woodcockem a jeho spoluautory započato úsilí několika generací o hledání lepšího porozumění podstatě lidských kognitivních schopností, jak lze tyto schopnosti v aplikovaném standardizovaném měření operacionalizovat a jak tyto schopnosti ovlivňují učení. Dnes toto úsilí pokračuje v podobě rozvíjející se základny znalostí známé jako moderní CHC teorie.

Hlavním cílem první revize testu Woodcock-Johnson testu byla formulace modelu lidské inteligence, kterou by bylo možné použít pro vývoj a interpretaci testu. Richard Woodcock se v roce 1985 zúčastnil konference pořádané University of Illinois na počest proslulého výzkumníka v oblasti inteligence, Lloyda Humphreyse. Na této konferenci prezentoval John Horn teorii lidské inteligence, která později vešla ve známost jako *Gf-Gc* teorie (Horn, 1991). Hornův vhled do struktury lidských intelektových schopností Woodcocka oslovil a nabízel se jako vhodný interpretační model pro *Woodcock-Johnson Psycho-Educational Battery-Revised* (Woodcock & Johnson, 1989). Vycházejí z Hornovy přednášky, Woodcock s pomocí Kevina McGrewa a za četných konzultací s Johnem Hornem a Jackem

⁴ V originále se píše o jazykových testech ve španělštině a angličtině.

Carrollem sloučil existující explorační a konfirmační faktorové analýzy WJ z roku 1977. Tyto analýzy vyústily v podrobný projekt pro plánování a organizaci revize WJ, aby se více přiblížila *Gf-Gc* teorii. Ve spolupráci s Nancy Mather a dalšími kolegy byly do nové verze *Woodcock-Johnson Psycho-Educational Battery-Revised* přidány nové testy, čímž se testová baterie stala předním modelem pro měření širokých kognitivních a školních schopností členěných dle *Gf-Gc* teorie (McGrew, 2005). Po vydání WJ-R obohatily Woodcockovy (1990) analýzy všech široce používaných inteligenčních baterií porozumění testům WJ-R, stejně jako schopnostem *Gf-Gc* měřeným jinými intelektovými bateriemi.

Podrobnější specifikace povahy a komplexnosti lidských kognitivních schopností se stala hlavním cílem při revizi *Woodcock-Johnson III* (Woodcock, McGrew, & Mather, 2001, 2007). Na tuto verzi měla největší vliv Carrollova publikace z roku 1993 *Human Cognitive Abilities: A Survey of Factor Analytic Studies*. Carrollova práce, často popisovaná jako teorie tří vrstev (Carroll, 1993, 1998, 2003), je postavena na výsledcích faktorových analýz 461 psychometrických souborů testových dat, které naznačují, že lidské kognitivní schopnosti mohou být rozčleněny do hierarchie o třech vrstvách. Carroll identifikoval přibližně 70 specifických nebo úzkých (vrstva I) kognitivních schopností, které měly vlastnosti Dobře replikovaného společného faktoru („*Well Replicated Common Factor*“ [WERCOF]) identifikovaným Hornem a jeho spolupracovníky (Ekstrom, French, & Harman, 1979; Horn, 1989). Carroll navíc seskupil úzké schopnosti do širších kategorií kognitivních schopností (vrstva II), které jsou podobné širokým faktorům *Gf-Gc* popsaným Hornem a jeho kolegy. Po přezkoumání dřívějších i současných teorií si Carroll všiml mnoha podobností a dospěl k závěru, že Hornův-Cattellův *Gf-Gc* model „patrně nabízí nejvíce odůvodněný a rozumný přístup k přijatelné teorii struktury kognitivních schopností“ (s. 62). Následně získala syntéza Cattellova-Hornova-Carollova výzkumu označení *CHC teorie* (McGrew, 2005, 2009) a byla využita jako východisko pro další rozšíření širokých faktorů WJ III. U většiny faktorů toho bylo dosaženo tak, že byly vytvořeny klastrové skóry širších schopností z dvou nebo více testů kvalitativně odlišných úzkých schopností (vrstva I). Carroll také popsal faktor obecné intelektové schopnosti (*g*), který je na vrcholu jeho modelu tří vrstev. Tak odůvodnil věcnou validitu diferenciálně váženého skóru obecné intelektové schopnosti (*g*) ve WJ III.

V roce 1994 nabídl John Carroll (1998) na Virginské univerzitě vlastní kritiku své teorie tří vrstev. Mezi jinými úvahami také upozornil na to, že specifikace v jeho teorii byly postaveny na značné subjektivitě ve výběru a třídění faktorů z nezávisle na sobě vybraných souborů dat. Navíc poznamenal, že jeho specifikace schopností vycházela primárně ze skórů psychometrických testů a pro validizaci navržených konstruktů jsou nezbytné další soubory dat získané pomocí jiných vědeckých metod. Obzvláště tyto dvě námitky stály v centru zájmu při vývoji WJ IV. *CHC* teorie se rozvíjela nad rámec své původní specifikace (Schneider & McGrew, 2012) prostřednictvím jak zjednodušení, tak další elaborace (viz Přílohu A). Navíc byly skrze vývoj WJ IV prováděny další studie zaměřené na ověření validity, modifikaci či větší objasnění některého z teoretických konstruktů navržených Cattellem, Hornem, Carrollem, Woodcockem a jejich kolegy.

Revidované nebo nové *CHC* široké a úzké schopnosti ve WJ IV

Interpretační model WJ IV odráží tu nejmodernější specifikaci *CHC* teorie, která existovala v době jeho publikování. Analýzy normativních vzorků WJ-R, WJ III a WJ IV poskytly tři rozsáhlé, multi-schopnostní soubory dat, které byly použity buď k potvrzení, nebo revizi výchozí specifikace konstruktů. Změna interpretačního pojetí byla podpořena dalšími poznatky z neurovědních výzkumů. Nejvýznamnější změny interpretačního modelu WJ IV se nachází v současném pojetí pracovní paměti, rychlosti lexikálního přístupu a paměti na zvukové vzorce.

Krátkodobá pracovní paměť (*Gwm*)

Ve WJ III pojem *krátkodobá paměť* (*Gsm*) odkazoval na schopnost vnímat a udržet v bezprostředním vědomí jednotky informace v okamžité situaci. Faktor *Gsm* ve WJ III byl odvozen primárně z publikací Horna (1965, 1988, 1989, 1991) a Horna a Nolla (1997), ačkoliv Carrollův (1993) faktor vrstvy II, Obecná paměť a učení (2Y) také obsahoval úlohy na rozsah auditivní paměti (*MS*). Navíc ve WJ III byly odlišně chápány úlohy, které staví čistě na zvukovém uložení a uchování informace, a úlohy, které vyžadují aktivní manipulaci s informací uloženou v bezprostředním vědomí. Tyto úlohy byly definovány jako míry pracovní paměti (*WM*).

Během desetiletí, které uplynulo od vydání WJ III, začal být termín *pracovní paměť* intenzivně citován v neurovědním výzkumu na základě operacionálních definic, které buď nahrazují, nebo zahrnují konstrukt krátkodobé paměti. Začalo být zřejmé, že umístění pracovní paměti v hierarchii pod *Gsm* v rámci CHC teorie je zastaralé a mylné (Dehn, 2008). Pracovní paměť ve svém nejširším smyslu odkazuje k dynamickému dočasnému úložnému systému, který dovoluje, aby informace – buď senzorické podněty, nebo předchozí znalosti – byly udržovány v bezprostředním vědomí a používány (Goldman-Rakic, 1992; Miller, Galanter, & Pribram, 1960). Na druhou stranu termín *krátkodobá paměť* znamená pro mnohé současné neurovědce výhradně „úlohy, které využívají významné úložiště, ale současně jen minimální zpracování“ (Gathercole & Alloway, 2008, s. 21). V odborném názvosloví *pracovní paměť* odkazuje k širšímu, komplexnějšímu konstrukt, oproti rozsahu krátkodobé paměti (WS). *Pracovní paměť* naznačuje aktivní manipulaci s informacemi a může pro dosažení cíle potenco- vat další kognitivní funkce, včetně vizualizace nebo paměti či procesů vyhledávání.

Ve WJ IV je krátkodobá pracovní paměť (*Gwm*) definována jako široká kognitivní schopnost, která v souladu s aktuálními neurovědními výzkumy představuje dynamický systém sloužící jak jako dočasné úložiště, tak k manipulaci s informacemi v lidské kognici. Oproti úzkému klastru ve WJ IV Rozsah auditivní paměti (MS), který je tvořen výhradně úlohami, které testují kapacitu v procesech mluveného jazyka za použití kódování založeného na zvucích, klaster Krátkodobá pracovní paměť (*Gwm*) ve WJ IV sestává z úloh, které zahrnují jak dočasné uložení, *tak* aktivní znovuvybavení nebo manipulace se slovy a čísly. V důsledku komplexnosti úlohy může být pro úspěšné zvládnutí úkolu v krátkodobé pracovní paměti častěji vyvolána vizuální představa nebo sémantická reprezentace. Klaster Krátkodobá pracovní paměť (*Gwm*) ve WJ IV může být také pojat jako míra kapacity krátko- době paměti (WM), diagnosticky důležitý a měřitelný aspekt širěji pojímaného konstrukt pracovní paměti používaného v současné neurovědě. Kapacita krátkodobé paměti je důležitým interpretačním konstruktem. Omezená kapacita pracovní paměti může působit jako překážka při učení, a tedy mít spojitost s poruchami učení (Gathercole, 2004; Gathercole, Lamont, & Alloway, 2006). Mnoho dětí s odhalenými poruchami v oblasti čtení, psaní nebo matematice vykazuje s tím související narušení pracovní paměti (Bull & Scerif, 2001; de Jong, 1998; Mayringer & Wimmer, 2000; Siegel & Ryan, 1989; Swanson, 1994; Swanson, Ashbaker, & Lee, 1996). Omezená kapacita pracovní paměti může také být kauzálním faktorem v poruchách učení, které nejsou doménově specifické. Gathercole a Alloway (2008) tvrdí: „Poruchy učení, které prostupují čtením a matematickými nebo jazykovými dovednostmi, se ukazují jako typické pro děti s oslabeným fungováním pracovní paměti“ (s. 25).

Paměť na zvukové vzorce (UM)

Carroll (1993) identifikoval úzkou schopnost auditivního zpracování, která se zdála zahrnovat paměť jakožto klíčovou složku. Přestože jeho datový soubor byl malý a sestával primárně z testů používajících vzorce neřečových zvuků, zaujala Carrollova možnost rozlišit auditivní faktor analogický s úzkým faktorem vizuální paměti (MV). V tradičním kognitivně-psychologickém výzkumu Neisser (1967) postuloval existenci echoické paměti neboli senzorického registru, jehož funkcí je retence auditivních informací po krátký časový úsek (3 až 4 vteřiny). Neisser, který je dnes považován za otce kognitivní psychologie, navrhnul, že jazyková funkce echoické paměti může spočívat v tom, že podrží dočasný auditivní vstup do doby, než je přidán další auditivní vstup, takže sekvence zvuků může být zpracována jako ucelená jednotka, jakou je celé slovo. O desítky let později jiný výzkumný směr, předtím nesouvisející s CHC teorií, představil schopnost fonologického uložení neboli funkci krátkodobé paměti pro řečové zvuky (Gathercole & Baddeley, 1989; Gathercole, Willis, Emslie, & Baddeley, 1992, 1994). Tyto a jim podobné studie ukazovaly na potřebu rozšířit CHC operacionální definici paměti na zvukové vzorce (UM) tak, aby zahrnovala řečové zvuky, což vedlo k zařazení Testu 12: Opakování pseudoslov do WJ IV COG.

Kognitivní komplexita

Jedním z interpretačních rysů, který je konstantní napříč všemi vydáními testu Woodcock-Johnson, je analýza testových požadavků prostřednictvím úrovně kognitivní komplexity. WJ, WJ-R i WJ-III byly cíleně navrženy tak, aby obsahovaly úkoly pokrývající kontinuum od jednoduchých kognitivních operací až po komplexní kognitivní procesy. Ve WJ IV bylo vynaloženo úsilí ke zvýšení požadované

kognitivní komplexity pro vybrané testy a klastry, aby bylo možné poskytnout větší ekologickou validitu a interpretační relevanci měření.

Vyšší kognitivní komplexity testu je obvykle dosaženo tím způsobem, že v testu je zkombinováno měření více než jedné úzké CHC schopnosti. Druhou možností je zvýšit kognitivní komplexitu nároků na zpracování informací *uvnitř* specifických úzkých CHC domén (Lohman & Lakin, 2011; McGrew, 2012). Tato druhá forma kognitivní komplexity, kterou je třeba nezaměňovat s faktorovou komplexitou, klade větší nároky na kognitivní zpracování informací (kognitivní zátěž), vyžaduje přesnější zacílení klíčových kognitivních zdrojů (pracovní paměti a kontroly pozornosti) a volá po zapojení větší kognitivní kontroly či exekutivních funkcí (Arend et al., 2003; Jensen, 2011; Lohman & Lakin, 2011; Marshalek, Lohman, & Snow, 1983). Tento druhý přístup zvyšování kognitivní komplexity testu byl výchozím principem při vývoji WJ IV.

Zvýšení kognitivní komplexity je zcela evidentní například ve složení klastru z WJ IV COG Auditivní zpracování⁵.

Auditivní zpracování (Ga)

Dva testy (Test 5: Fonologické zpracování a Test 12: Opakování pseudoslov), které ve WJ IV COG tvoří klaster Auditivního zpracování, byly vyvinuty pro měření kognitivně komplexních, ekologicky relevantních procesů, které zahrnují schopnosti auditivního zpracování. Na základě Carrollových publikací (1993) a WJ III (Woodcock, McGrew, & Mather, 2001, 2007) vznikla záplava výzkumů auditivních schopností (Schneider & McGrew, 2012). Široký soubor charakteristik auditivního zpracování (Ga) byl dáván do souvislosti s poruchami čtení, psaní a jazyka. Schopnosti Ga jsou nyní považovány za klíčové stavební kameny při vývoji jazyka a obecných kognitivních schopností (Conway, Pisoni, & Kronenberger, 2009). Do WJ IV byly pro měření některých těchto důležitých schopností přidány dva nové testy auditivního zpracování. Každý z testů je založen na kombinaci úzkých schopností, které pokrývají jednu nebo více jiných širokých schopností. Těmito dvěma novými testy jsou Test 5: Fonologické zpracování a Test 12: Opakování pseudoslov⁶.

Test 5 Fonologické zpracování měří tři aspekty hláskového zpracování řeči, které vedou k vytvoření na hláskách založených lexikálních reprezentací. Tyto typy úloh, někdy též označovány jako fonologické uvědomění (angl. *phonological awareness*) či fonologická senzitivita (angl. *phonological sensitivity*) (Pufpaff, 2009), úzce souvisí s učením slov (de Jong, Seveke, & van Veen, 2000). Fonologické zpracování je považováno za důležité během učení se čtení, protože fonologie se během vyslovování opírá o ortografii (Liberman, Shankweiler, & Liberman, 1989; Wagner, Torgesen, Laughton, Simmons, & Rashotte, 1993; Wagner, Torgesen, & Rashotte, 1994). Subtest 5A Vyhledávání slov zjišťuje schopnost nalézt podle fonologické nápovědy v mentálním lexikonu příslušné slovo. Nejjednodušší položky z Vyhledávání slov vyžadují, aby testovaná osoba ukázala na obrázek, který začíná na stejnou hlásku jako podnětové slovo. Ostatní položky vyžadují nalézt slovo, které má příslušný fonetický prvek na příslušném místě. Subtest 5B Slovní fluence zjišťuje rozsah, do jakého může být prostřednictvím fonologického spojení aktivována slovní síť. Subtest 5C Substituce měří typ fonologické senzitivity, díky níž je možné v pracovní paměti nahrazením hlásky nebo části slova vytvořit nové slovo. V kontextu moderní CHC teorie celkový testový skóre odráží kombinaci schopností auditivního zpracování, fluence a vybavování z dlouhodobé paměti.

Test 12 Opakování pseudoslov je kognitivně komplexní test, který měří kombinované schopnosti auditivního zpracování a krátkodobé pracovní paměti, někdy označované jako fonologická krátkodobá paměť. Úloha vyžaduje dočasné uložení fonologických úseků v bezprostředním vědomí, což je schopnost, ve které existují výrazné individuální rozdíly mezi lidmi (Gathercole & Baddeley, 1989; Gathercole et al., 1992, 1994). Nedávné studie s úlohami na opakování pseudoslov naznačují spojitost mezi opakováním pseudoslov a učením se novým slovům (Gathercole, 2006; Michas & Henry, 1994) a možnost, že testy opakování pseudoslov mohou pomoci nahlédnout do vývoje slovní zásoby (Edwards, Beckman, & Munson, 2004). Navíc byl výkon v úlohách s opakováním pseudoslov ve vztahu s poruchami čtení a jazyka (Archibald & Gathercole, 2006; de Bree, Wijnen, & Gerrits, 2010; Gathercole, 2006)⁷.

5 Další příklad se týká testu z baterie školních dovedností, která do češtiny převedena nebyla.

6 Další úkoly vyžadující fonetické zpracování jsou součástí baterie testů mluveného jazyka, které do češtiny převedeny nebyly.

7 Zde byla vynechána pasáž týkající se plynulosti čtení z baterie školních dovedností.

Organizace do tří odlišných baterií

Celý WJ IV je rozdělen do tří odlišných baterií tak, aby usnadnil široký rozsah na míru šitých a obsáhlých vyšetření prováděných jedním či více odborníky. V závislosti na účelu vyšetření mohou být baterie používány buď samostatně, nebo v kombinaci s testy či klastry z jedné nebo obou dalších baterií. Celý systém testů a klastrů WJ IV je navržen tak, aby umožnil široké pokrytí individuálně administrovaných vyšetření nejdůležitějších schopností v různých variantách sezení a časového harmonogramu. Velký věkový rozsah i širší pokrytí umožňují, aby testy a klastry byly použity pro školní, klinické nebo výzkumné potřeby, a to od předškolních dětí po osoby vysokého věku⁸. Procedury použité při vývoji všech klastrů WJ IV a postupy srovnávání popsané v této kapitole jsou prezentovány v dalších kapitolách tohoto manuálu.

WJ IV Testy kognitivních schopností

WJ IV COG obsahuje 18 testů pro měření obecné inteligence, širokých a úzkých kognitivních schopností, školních doménově specifických vloh a přidružených aspektů kognitivního fungování. Dvě na stojánku umístěné testové knihy se nazývají Standardní baterie (Testy 1–10) a Rozšířená baterie (Testy 11–18). Tabulka 1-1 poskytuje informace o interpretační organizaci testů WJ IV COG. Některé testy měří samostatnou, úzkou kognitivní schopnost tak, jak je definována v aktuálním pojetí CHC teorie. Další testy jsou vytvořeny nebo revidovány tak, aby zdůrazňovaly kognitivní komplexitu. Při vybírání testů do finální verze baterie byl kladen důraz na psychometrické charakteristiky a důležitost schopnosti nebo kombinace schopností pro učení, intervence nebo přizpůsobení.

Testy byly řazeny tak, aby byla maximalizována jejich interpretační hodnota s co nejnižšími náklady na testování a aby byla vytvořena základna pro výběr dalších testů. Deset testů, které tvoří Standardní baterii, mají představovat standardní vyšetření kognitivních funkcí pro různé účely vyšetření:

- Testy 1 až 3⁹ tvoří skór Zkrácené inteligenční škály (BIA), sestávající z jednoho zástupce testu pro porozumění – znalosti (*Gc*), fluidní inteligenci (*Gf*) a krátkodobou pracovní paměť (*Gwm*). Tato podmnožina z celé kognitivní baterie WJ IV velmi silně koreluje s obecnou inteligencí a (na rozdíl od skóru BIA z WJ III) nezahrnuje míru rychlosti kognitivního zpracování (*Gs*).
- Testy 1 až 7 tvoří skór obecné inteligence (*GIA*), kdy každý test zastupuje jednu z CHC domén schopností, a to porozumění – znalosti (*Gc*), fluidní inteligenci (*Gf*), krátkodobou pracovní paměť (*Gwm*), rychlost kognitivního zpracování (*Gs*), auditivní zpracování (*Ga*), dlouhodobou paměť (*Glr*) a vizuální zpracování (*Gv*). Každý test zahrnutý mezi prvních sedm testů (a) byl silným indikátorem příslušné CHC domény schopností, (b) měl vysoké náboje na faktor obecné inteligence (*g*), (c) měl vysokou kognitivní komplexitu a (d) byl mezi nejsilnějšími prediktory školního výkonu ve WJ IV ACH. Tato sada představuje požadovaný soubor testů pro vyšetření silných a slabých stránek mezi testy a klastry, které měří další kognitivní schopnosti.
- Testy 1 až 10 obsahují všechny klastry a funkce zmíněné výše a k tomu navíc také klastry porozumění – znalosti (*Gc*), fluidní inteligenci (*Gf*), krátkodobou pracovní paměť (*Gwm*), kognitivní efektivitu (*CE*) a kombinovaný skór *Gf-Gc*.

Osm testů, které obsahuje Rozšířená baterie, zlepšují interpretační potenciál informací zjištěných pomocí Standardní baterie. Jakýkoliv z těchto testů může být administrován selektivně a nezávisle, nebo být použit pro získání skóru klastru pro některou z následujících širokých a úzkých kognitivních schopností: rychlost kognitivního zpracování (*Gs*), auditivní zpracování (*Ga*), dlouhodobá paměť¹⁰ (*Glr*), vizuální zpracování (*Gv*), rychlost vnímání (*P*), kvantitativní uvažování (*RQ*) a početní dovednosti (*N*). Kombinujeme-li výsledky s testy z WJ IV OL¹¹, můžeme získat klastry úzkých kognitivních schopností slovní zásoby (*LD/VL*) a rozsahu auditivní paměti (*MS*). Byly-li administrovány testy 1 až 7, jakýkoliv test nebo klástr navíc, který je administrován, může být zařazen do procedury výpočtu intra-kognitivních variací pro zhodnocení relativních slabých nebo silných stránek.

8 Českým examinátorům je zatím k dispozici pouze první baterie – Woodcock-Johnson IV Testy kognitivních schopností (WJ IV COG), která je stručně představena v následující kapitole.

9 V tomto a následujících odstavcích jsou popisovány již dříve zmíněné klastry, tzn. kompozitní skóry získané z kombinací vhodných testů WJ IV.

10 Tato schopnost bývá také nazývána „dlouhodobá paměť a vybavení“ nebo „fluence vybavování“ (viz Příloha A).

11 Zkratka OL znamená Oral Language, což je baterie, která v české adaptaci WJ IV obsažena není.

Tabulka 1-1.
WJ IV COG Tabulka
selektivního testování
ukazující testy
a interpretační klastry

			Obecná inteligence (GfA) Zkrácená inteligenční škála (BfA) Kombinovaný skóre Gf-Gc Porozumění – Znalosti (Gc) Fluidní inteligence (Gf) Krátkodobá paměť (Gwm) Rychlost kognitivního zpracování (Gs) Dlouhodobá paměť (Ga) Vizuální zpracování (Gv) Kvantitativní uvažování (Hq) Rychlost vnímání (In) Kognitivní efektivita (Ce)													
Standardní baterie	Test 1	Slovník	■	■	■	■										
	Test 2	Číselné řady	■	■	■	■								■		
	Test 3	Verbální pozornost	■	■				■								□
	Test 4	Hledání písmen	■					■							■	■
	Test 5	Fonologické zpracování	■						■							
	Test 6	Reprodukce příběhů	■							■						
	Test 7	Vizualizace	■								■					
	Test 8	Všeobecné znalosti			■	■										
	Test 9	Formování konceptů			■	■										
	Test 10	Obrácené číselné řady						■						■		■
Rozšířená baterie	Test 11	Hledání čísel												■	■	□
	Test 12	Opakování pseudoslov								■						
	Test 13	Audio-vizuální učení									■					
	Test 14	Rozpoznávání obrázků										■				
	Test 15	Analýza - Syntéza					□						■			
	Test 16	Řazení názvů a čísel						□								
	Test 17	Vyhledávání dvojic							■							
	Test 18	Paměť na slova														

- Testy nezbytné pro vybraný klaster ze seznamu.
□ Doplnkové testy nezbytné pro rozšířený klaster ze seznamu.

Návrhy použití a interpretace testových skóre WJ IV

V souladu se *Standards pro pedagogické a psychologické testování* (Standards for Educational and Psychological Testing; American Educational Research Association [AERA], American Psychological Association [APA], National Council of Measurement in Education [NCME], 1999, 2001) by argumenty ve prospěch validity testových skóre měly být přinášeny během procesu vývoje testu, společně s jasným vyjádřením návrhů podporujících doporučená použití a interpretaci skóre. Následně „Proces validizace postupuje vyslovením těchto tvrzení a shromažďováním důkazů na jejich podporu.“ (s. 18). *Standards pro pedagogické a psychologické testování* obsahují doporučení nejlepší praxe pro návrh testu, vývoj a interpretaci. Dodržení těchto standardů zajišťuje, že vývojáři testu postupy navrhování i vývoje testu pečlivě promýšleli a jsou si vědomi toho, jak tyto postupy ovlivňují validitu interpretací vycházejících z testových skóre. Během návrhu a vývoje WJ IV byl kladen velký důraz na to, aby všechny relevantní standardy byly dodrženy. Díky vedení dokumentace o vývojovém procesu a technických charakteristikách testu mohou uživatelé posoudit, v jakém rozsahu jsou získané skóre validní pro jejich použití a interpretaci. Přestože soulad se *Standards* nezaručuje validitu testových skóre, je uživateli prospěšný tím, že mu poskytuje rámec pro zhodnocení předložených důkazů o validitě.

Podpora pro užití a interpretaci skóre WJ IV staví na důkazech o validitě shromážděných ve třech předešlých řadách testu (WJ, WJ-R a WJ-III) i výstupech velkého počtu studií souvisejících s CHC teorií. Další důkazy o validitě obsahuje dokumentace východisek, cílů a teoretického pozadí revizí a postupů užitých při vývoji testu, standardizaci a procesu vytváření norem. Podpora užití a interpretace testových skóre WJ IV také vychází z údajů o technických charakteristikách testu, včetně jeho vnitřní struktury a vztahů skóre WJ IV s dalšími proměnnými. *Technický manuál Woodcock-Johnson IV* (McGrew, LaForte, & Schrank, 2014) v celém svém rozsahu slouží jako důkazní materiál pro podporu validity skóre WJ IV měřících kognitivní schopnosti, jazykové schopnosti a školní dovednosti jedince.

Design testu a postup vývoje

Vývoj testu *Woodcock-Johnson IV* (WJ IV) (Schrank, McGrew a Mather, 2014a) postupoval dle tradičních stadií a procedur, včetně podrobné revize a aktualizace testových specifikací, tvorby nových testů a položek, ověření a psychometrického zhodnocení nových položek, škálování skupin položek, normování a výběru výsledných testových forem. *Standardy pro pedagogické a psychologické testování* (Americká asociace výzkumu vzdělávání (AERA), Americká psychologická asociace (APA) a Národní rada pro měření ve vzdělávání (NCME), 1999, 2001) obsahují řadu doporučení pro nejlepší praxi při tvorbě testů a jejich revizí. Tato kapitola slouží jako doklad toho, že relevantní doporučení byla během designu testu WJ IV a jednotlivých fází vývoje prostudována a uplatněna.

Aktualizace testových specifikací

WJ IV je postaven na teoretickém, strukturním a interpretačním rámci testu *Woodcock-Johnson III* (WJ III) (Woodcock, McGrew a Mather, 2001, 2007). Cíle revizí a design testu WJ IV byly založeny na soudobé teorii CHC a výzkumu, identifikaci potřeb odborné praxe a zpětné vazbě od uživatelů WJ III. Tyto cíle jsou popsány v první kapitole tohoto manuálu. Cíle revize ovlivnily aktualizaci specifikace testů, včetně tvorby několika nových testů a interpretačních klastrů a rozšíření několika existujících testů a interpretačních klastrů z WJ III. Některé z nových testů ve WJ IV byly vyvinuty se záměrem zvýšit ekologickou a prediktivní validitu interpretací klastrů, ostatní mají za cíl rozšířit nástroje k měření širokých nebo úzkých schopností identifikovaných v nedávném výzkumu CHC.

Vývoj nových testů WJ IV

Do WJ IV bylo přidáno osm nových testů: Verbální pozornost, Hledání písmen, Fonologické zpracování, Opakování pseudoslov, Segmentace, Ústní čtení, Vybavování čtení a Plynulost čtení slov. Následuje stručné zdůvodnění, cíle vývoje a typy položek pro jednotlivé testy¹².

- *Verbální pozornost.* Tento test byl vyvinut jako ekologicky validní nástroj k měření pracovní paměti. Cílem bylo vyvinout položky podobné úkolům, které jsou po studentech požadovány ve výuce či po dospělých v typickém pracovním prostředí. Podnět ve Verbální pozornosti je podobný položkám z testu Řazení objektů a čísel (Sluchová pracovní paměť ve WJ III). Testovaný si nejprve poslechne verbálně prezentovanou řadu obsahující jména zvířat promíchaná s číslicemi. Namísto opakování řady nebo třídění položek na zvířata a číslice (jak je požadováno v testu Řazení objektů a čísel), zde musí proband řadu podržet v paměti a poté zodpovědět konkrétní otázku, která se k řadě vztahuje. Testovaná otázka dopředu nezná do okamžiku, kdy jsou všechny položky přečteny a jsou mu sděleny instrukce specifické pro položku. Podněty v položkách byly vytvořeny tak, aby tvořily škálu od velmi snadných (např. řada jednoho zvířete a jedné číslice, načež se examinátor dotáže na jméno zvířete) až po velmi komplexní (např. seznam devíti jmen zvířat a mezi ně zařazených číslic, načež se examinátor ptá na specifickou část řady).

¹² Následující odstavce se týkají pouze testů z baterie Testy kognitivních schopností.

- *Hledání písmen.* Tento test byl vyvinut k měření percepční rychlosti a ortografického zpracování. Výzkum ortografického zpracování a ortografických znalostí (Apel, 2011; Barker, Torgesen a Wagner, 1992; Berninger, 1990; Berninger et al., 2006) naznačuje, že někteří zdatní čtenáři rozeznávají běžné ortografické vzorce bez problémů, zatímco jiní lidé, kteří mají s rychlým čtením problémy, těžko rozpoznávají běžné vzorce písmen. V tomto testu bylo cílem vyvinout položky, které by rozlišovaly mezi lidmi, kteří rychle a snadno rozeznávají ortografické vzorce, a těmi, komu se to tak efektivně nedaří. Formát položek je založen na úloze z podobného Gs testu, Hledání čísel (Vizuální přiřazování z WJ III), který měří rychlost, s jakou proband dokáže vizuálně rozlišit mezi čísly resp. symboly. V tomto novém testu Hledání písmen vyžaduje úloha, aby proband našel a zakroužkoval dvě odpovídající sekvence písmen v řádku, který obsahuje několik kombinací distraktorů. Cílová (odpovídající) sekvence písmen je složena z běžných českých vzorců písmen, např. st nebo er. Distraktory jsou nemožné nebo méně časté kombinace jako nh nebo bg¹³. Položky z Hledání písmen sahají od jednoduchých, s cílovými „sekvencemi“ a s distraktory o délce jediného písmene, po obtížnější se čtyřpísmennými vzorci
- *Fonologické zpracování.* Tento test, složený ze tří subtestů, je vytvořen tak, aby nabídl kognitivně komplexní (viz definice v první kapitole) nástroj k měření fonetického kódování nebo schopností fonologického zpracování. Subtest C, Substituce, byl zařazen v testu Zvukového povědomí ve WJ III. Subtest A, Vyhledávání slov, a subtest B, Slovní fluence, byly vyvinuty pro WJ IV.
 - *Fonologické zpracování - Vyhledávání slov.* Tento subtest byl vyvinut jako ekologicky validní test zaměřený na kognitivní procesy nutné k přístupu ke slovům na základě fonologie. Tento nový test je podobný testu Spojování zvuků, jinému testu ve WJ IV. V testu Spojování zvuků proband poslouchá sérii slabik nebo fonémů, načež je požádán o spojení zvuků do slova. Ve WJ III bylo Spojování zvuků zařazeno do klastru GIA. Zpětná vazba od uživatelů naznačila, že explicitní instrukce ke spojování fonémů do slov může vést k efektu zácviky; děti systematicky učené rozpoznávat zvuky a poté je spojovat do slov měly tendence zlepšovat svoje skóre ve Spojování zvuků. Z tohoto důvodu, a také s cílem naplnit cíl zařazení kognitivně komplexních testů Ga do *Woodcock-Johnson IV Testy kognitivních schopností* (WJ IV COG) (Schränk, McGrew a Mather, 2014b), bylo rozhodnuto o přesunutí Spojování zvuků do *Woodcock-Johnson IV Testy mluveného jazyka* (WJ IV OL) (Schränk, Mather a McGrew, 2014b) jako doplňku k testu Segmentace. Cílem nového subtestu Vyhledávání slov bylo zahrnout položky o široké škále obtížností, které nebudou náchylné k efektu zácviky. Nejsnazší položky z Vyhledávání slov po probandovi vyžadují, aby ukázal na obrázek, který začíná na stejný zvuk jako slovo nabídnuté administrátorem. Zbylé položky vyžadují, aby si proband vybavil takové slovo ze své slovní zásoby, které na určitém místě (např. na začátku) obsahuje určitý fonemický element. Většina položek je administrována pomocí audionahrávky.
 - *Fonologické zpracování - Slovní fluence.* Tento subtest byl vytvořen za účelem zvýšení kognitivní komplexity testu Fonologické zpracování, klastru Zvukového zpracování (Ga) a klastru GIA. V tomto subtestu má testovaný za jednu minutu jmenovat co nejvíce slov, která začínají určenou hláskou.
 - *Fonologické zpracování - Substituce.* Tento subtest byl subtestem v testu Zvukové uvědomění ve WJ III. Do WJ IV je zahrnut pro zvýšení kognitivní komplexity testu Fonologické zpracování. V tomto subtestu má zkoušený nahradit část slova tak, aby vytvořil slovo nové. Takto proband například obdrží slovo jíst a instrukci, aby zaměnil j za p tak, aby vytvořil nové slovo (píst)¹⁴. Většina položek je administrována pomocí audionahrávky.
- *Opakování pseudoslov.* K přidání tohoto testu došlo v reakci na soudobý výzkum, který naznačuje, že schopnost podržet zvuky řeči ve fonologické paměti souvisí s učením nových slov a že tento typ úloh je senzitivní vůči individuálním rozdílům a poruchám v oblasti čtení

13 Zde byl překlad upraven tak, aby se rovnou týkal české verze testu Hledání písmen. V originále byly příklady shodujících se skupin oa a sh, příklady nemožných nebo málo frekvencovaných ao a hs.

14 I zde jsme překlad doplnili příkladem z české verze subtestu. V originále bylo podnětovým slovem taller, ve kterém se má tall zaměnit za small, aby vzniklo nové slovo smaller.

a jazykového vývoje (Archibald a Gathercole, 2006; Baird, Slonims, Simonoff a Dworzynski, 2011; Gathercole a Baddeley, 1989; Gathercole, Willis, Emslie, a Baddeley, 1992, 1994; Nation a Hulme, 2010). Cílem bylo vyvinout náročný nástroj k měření krátkodobé fonologické paměti. Formát položek je založen na podobném nástroji (Zvukové mimikry) z Goldman–Fristoe–Woodcock Auditory Skills Battery¹⁵ (Goldman, Fristoe a Woodcock, 1974). Nové položky byly vyvinuty tak, aby obsahovaly pseudoslova s fonematickými vzorci běžnými v češtině¹⁶. Zkoušený si nejprve poslechne pseudoslovo a poté je má zopakovat. Položky jsou od velmi jednoduchých (jedna slabika) po velmi obtížné (sedm a více slabik). Nejsnazší jednoslabičné položky jsou zadávány administrátorem, zbylé položky audionahrávkou.

Každá z položek v nových testech WJ IV prošla extenzivní pilotáží. Po tvorbě množiny položek proběhla administrace omezenému vzorku, aby se prověřil formát položek a potvrdila se jasnost příslušných instrukcí. Po provedení všech nezbytných úprav byl poté každý test administrován příležitostně sebranému vzorku přibližně 100 až 200 probandů s širokým spektrem věků a schopností. Smyslem této fáze pilotáže byla získání předběžných odhadů obtížnosti položek a jiných statistik tak, aby bylo možné posoudit, zda je před studií nutno provést další změny či vývoj.

Vývoj české verze¹⁷

Česká adaptace originální verze WJ IV COG, vydaná v USA v anglickém jazyce, si vzhledem k odlišnému jazykovému i kulturnímu prostředí vyžádala specifické úpravy. Největšími úpravami prošly testy pracující s verbálním materiálem, tedy Test 1 Slovník, Test 3 Verbální pozornost, Test 5 Fonologické zpracování, Test 6 Reprodukce příběhů, Test 8 Všeobecné znalosti, Test 12 Opakování pseudoslov, Test 13 Audio-vizuální učení, Test 16 Řazení názvů a čísel a Test 18 Paměť na slova. Postup adaptace sestával jednak z překladu originálních položek a následného posouzení relevance z hlediska jazykové i kulturní adekvátnosti a také z vytvoření zcela nových položek na stejném principu pro zajištění rovnoměrného rozložení jejich obtížnosti ve výsledném testu. Relevance položek byla ověřena v pilotní studii na českém vzorku a poté standardizační studií. Na základě výsledků statistických analýz (zejména analýz Raschova modelu) i kvalitativního posouzení získaných odpovědí byly pro jednotlivé testy vytvořeny definitivní seznamy položek v pořadí odpovídajícím vzestupné obtížnosti, případně u Testu 1 a 8 doplněny vhodné správné odpovědi či takové, které vyžadují dodatečný dotaz. Nástroj, který byl využit při tvorbě české verze, byl Český národní korpus (ČNK).

Test 1: Slovník

Při adaptaci Testu 1 byly v prvním kroku doslovně přeloženy položky z originálu, včetně správných odpovědí. Několik těchto položek nebylo vůbec zařazeno, a to zejména z důvodu absence jiných synonym v češtině (např. *chef* – kuchař) či blízkosti významů s jinými již použitými položkami. Seznam byl doplněn o několik položek ze subtestu Synonyma, který je součástí Krátkého inteligenčního testu (Říčan & Laciga, 2015) a několik bylo zcela nově vytvořeno autory adaptace. V souladu s originální verzí byly v seznamu ponechány různé typy synonym a antonym (např. položky tvoří nejen opozita vlastní, ale i konverzní a komplementární).

Relevantní správné odpovědi převzaté z originální verze byly ponechány, nerelevantní vyloučeny. Další správné odpovědi, které jsou typické pro český jazyk, byly doplněny na základě informací ze Slovníku českých synonym (Pala & Všianský, 2000). Ze specifík češtiny bylo nezbytné u sloves brát v potaz vid, který může měnit význam, a tak byla preferována shoda vidu výrazu v zadání položky a správné odpovědi (budovat – bourat vs. zbourat; čistit – špinit vs. zašpinit). Vzhledem ke specifickému tvoření záporu v češtině připojením negačního prefixu *ne-* bylo dále nutné nastavit podmínky tak, aby mezi probandovými odpověďmi nemohl takový způsob odpovídání převládat, jelikož poté by mohlo být obtížné poznat, zda proband skutečně má adekvátní slovní zásobu, či jen užívá nějaký naučený automatický postup řešení. U položek WJ IV shodných s WJ IE byly přejaty některé odpov-

¹⁵ Baterie auditivních dovedností Goldman-Fristoe-Woodcock nemá českou adaptaci.

¹⁶ V původní verzi se pochopitelně píše o angličtině.

¹⁷ Autorkou této části je Monika Vichová.

vědi z této verze. Na základě doslovného zápisu odpovědí z pilotní a standardizační studie byly poté navíc doplněny frekventované odpovědi, které byly jazykově relevantní, případně byly zařazeny k dodatečnému dotazu. Při vytváření seznamu byly voleny odpovědi nejbližší k nejtypičtějšímu významu slova a také adekvátní odpovědi frekventované v datech ze standardizační verze. Některé odpovědi probandů se mohou vázat k širšímu či přenesenému významu výrazu. Takové odpovědi nejsou vyloženě nesprávné, ale také nejsou zcela typické či konvenční. Tyto odpovědi proto byly opatřeny dodatečným dotazem, který ověří, zda proband zná i typičtější odpověď. Seznam odpovědí není autory považován za úplný a setká-li se examinátor s odpovědí, která není součástí seznamu, je na jeho posouzení, zda ji bude hodnotit jako správnou.

Test 3: Verbální pozornost

Při adaptaci Testu 3 do češtiny nebyly nutné žádné větší úpravy. Originální položky byly doslova přeloženy a dle stejného principu bylo autory testu vytvořeno několik nových položek navíc. Všechna zvířata v originální verzi mohou být známa také osobám žijícím v České republice, proto nebyla nutná žádná obměna. Finální pořadí položek v české verzi vzniklo na základě analýzy dat z pilotní a standardizační studie.

Test 5A: Fonologické zpracování – Vyhledávání slov

U Testu 5A jsme se opět drželi principu testu, ale jeho obsah bylo nutné přizpůsobit českému jazyku. První obrázková část byla přizpůsobena české slovní zásobě, z důvodu čehož bylo nutné vytvořit nové obrázky, které jsou použity v české verzi testu. V originální verzi sestává položka ze čtyř obrázků – jednoho podnětu (např. *dog*), jedné správné odpovědi (*duck*), jednoho fonetického distraktoru, tj. slova, které obsahuje stejnou hlásku, ale ne na první pozici (např. *frog*) a jednoho sémantického distraktoru, objektu, který má významovou nebo kontextovou spojitost s podnětem či jiným obrázkem v zadání (v originální verzi to byl obrázek kočky – *cat*). Tento princip byl zachován i při tvorbě položek pro českou verzi.

Ve druhé části 5A pro českou verzi byly vybrány takové hlásky či skupiny hlásek, jež se vyskytují ve slovní zásobě českých mluvčích, přičemž se stejně jako v originále může jednat o jakýkoliv slovní druh, jak obecná jména, tak propria, jak slova domácí, tak výpůjčky, nebo i cizí názvy známých míst či osob, které jsou v češtině používány v původním tvaru. Jako pomocník k ověření relevance vytvořených položek byl použit ČNK SYN2015 (Křen et al., 2015; Machálek, 2014). Opět bylo cílem vytvořit položky různé obtížnosti, od velmi snadných po velmi obtížné. Odhadovaná obtížnost byla stanovena dle dvou kritérií, a to množství různých výrazů v české slovní zásobě a jejich frekvence (nejsnadnější jsou hlásky, jež se vyskytují v mnoha slovech vysoké frekvence; o něco obtížnější jsou skupiny hlásek, které jsou zastoupeny v málo slovech, která jsou ale vysoce frekventovaná, či ve větším množství slov s nízkou frekvencí; nejvyšší obtížnost pak mají skupiny hlásek, které se vyskytují v malém množství slov, která jsou navíc velmi málo frekventovaná). Skutečná obtížnost byla ověřena pomocí odhadu parametrů Raschova modelu dat získaných ze standardizační studie. Jelikož jde o test fonologického uvědomění, při hodnocení správné odpovědi je klíčová zvuková stránka slova. U výrazů s hláskami, které podléhají asimilaci znělosti, však z důvodu zjednodušení skórování a zaručení větší shody mezi různými hodnotiteli doporučujeme hodnotit jako správné i takové odpovědi, kde v souladu s instrukcí je třeba pouze grafická stránka výrazu (např. ztráta jako odpověď slova začínajícího hláskou *z*, přestože foneticky je vyslovováno *s*). Je nutné zohlednit také specifickou grafického zápisu některých hlásek (např. výrazy *dívka*, *děti* apod. graficky můžou implikovat přítomnost hlásky *d*, ale z hlediska fonetického jde o hlásku *ď*).

Test 5B: Fonologické zpracování – Slovní fluence

V testu 5B autoři české adaptace původně zachovali hlásky použité v originální verzi, tedy *m* a *d*, avšak na základě výsledků pilotní studie se ukázala nutnost úpravy. Původně zvolená hláška *d* totiž v češtině má stejný grafický záznam jako hláška *ď*. Některými probandy nebyla tato odlišnost vnímána a mezi examinátory nepanovala shoda v tom, zda uznat odpovědi typu „dítě“ či „dívka“, které foneticky nezačínají na hlásku *d*, nýbrž *ď*. Nakonec byla vybrána hláška *s*, u níž sice také může

docházet k diskuzím (vlivem asimilace znělosti i gramatických pravidel může proband zaměňovat výrazy začínající na hlásku s a z), nicméně jde o hlásku, která se v češtině používá i v jiných adaptacích této úlohy a z hlediska obtížnosti je považována za neutrální (Nikolai et al., 2015; Štorková, Preiss, & Kopeček, 2004).

Test 5C: Fonologické zpracování – Substituce

V Testu 5C bylo nezbytné vytvořit všechny položky dle slovní zásoby češtiny. Ve shodě s originální verzí zněla instrukce u některých položek (zejména v první části testu) nahradit slovotvorný základ (např. v *Penny* – pen za sun) a u jiných nahradit jednotlivé hlásky či skupiny hlásek (p v *sip* za t nebo sm ve *small* za st). Všechny výchozí i cílové výrazy patří k běžné slovní zásobě češtiny. Česká verze na rozdíl od americké verze obsahuje dvě vlastní položky, které v instrukci kromě pouhé výměny hlásek žádají i odstranění jiných hlásek. Tyto položky byly zamýšleny jako nejobtížnější položky a z toho důvody byly také zařazeny do finální české verze.

Test 6: Reprodukce příběhů

Převod podnětového verbálního materiálu v Testu 6 znamenal zejména úpravu některých slovních spojení, slovosledu i přizpůsobení českému kulturnímu prostředí (počestění jmen osob či názvů míst). Cílem bylo v co nejvyšší míře zachovat logiku a smysl příběhu tak, jak je v originálu. Při překladu byla dána přednost známým a tedy frekventovaným slovům ve slovní zásobě češtiny, vzhledem k zacílení testu také na dětskou populaci byl pro ověření použit ČNK SCHOLA2010 (Šebesta, Golánová, Křen, & Procházka, 2010).

Některé položky si vyžádaly změnu pořadí slov či vět, aby byl výsledek v souladu s pravidly českého jazyka. Některé anglické jednoslovné výrazy nemají v češtině jednoslovnou variantu a bylo nutné je nahradit dvouslovným názvem či spojením více slov, nebo naopak v angličtině dvouslovná pojmenování lze přeložit jedním slovem. Tento způsob úpravy naštěstí nebyl problém, jelikož testová předloha nepracuje pouze se slovy, ale s celými prvky (označeny lomítkem /), takže jedinec pro získání bodu musí uvést všechny klíčové výrazy z celého prvku (v testu zaznačeny tučným fontem). Některé klíčové výrazy bylo možné přeložit jako víceslovná spojení, která jsou v češtině ustálená (např. *accelerator* – plynový pedál), spoluvýskyt jiných je spíše nahodilý (např. *hard to teach* – těžké něčemu naučit).

Rozmanitost výrazů použitých v české verzi kopírovala různost v originále. Cílem této snahy bylo docílit toho, aby se stejné slovo v jednom příběhu zbytečně neopakovalo, jelikož by to mohlo facilitovat zapamatování, což však u některých výrazů bylo obtížné nebo dokonce nemožné. Stejně jako v originále se uznává, použije-li proband v odpovědi výraz, který je synonymem ke slovu v přečteném příběhu. Doporučujeme uznávat obměny tvaru (zdrobnělina, změna rodu apod.) a dále takové výrazy, které jsou uvedeny ve Slovníku českých synonym (Pala & Všianský, 2000). V některých případech, kdy je (stejně jako v originále) cíleně použito v jednom příběhu pro vyjádření stejného objektu jiné slovo (např. policisté – strážníci), autoři doporučují pro udělení více bodů požadovat více různých označení (tedy nedávat dva body, řekne-li proband ve svém zopakování příběhu pouze dvakrát policisté, nebo dvakrát strážníci).

Z hlediska kulturních rozdílů bylo v české verzi nutné vypořádat se s produktem *candy bubbles*, který tvoří klíčový prvek jednoho z příběhů, avšak dostupnost tohoto produktu na českém trhu, a tedy oficiální české označení, nejsou autorům české verze známy.

Test 8: Všeobecné znalosti

V Testu 8 byly přeloženy originální položky, včetně správných odpovědí, a byla zhodnocena jejich relevance. Všechny položky v originální verzi obsahovaly objekty, s nimiž se může běžně setkat i jedinec žijící v České republice, proto nebylo nutné žádnou z nich vyřazovat pro neexistenci v českém prostředí. U některých slov byl po překladu význam čitelný z jednotlivých částí slova – např. měřidlo (*metre stick*) či krokoměr (*pedometer*), proto pro zajištění různé obtížnosti bylo několik takových položek nahrazeno výrazy podobného smyslu, ale s méně průhlednou strukturou (šuplera, tachometr). Další položky byly dotvořeny autory testu tak, aby šlo o v českém prostředí existující a zároveň

nepříliš oborově vyhraněné objekty. V 8B došlo k úpravě úvodní otázky – otázku “*What would people usually do with...*” jsme přeložili volněji jako „K čemu se obvykle používá...?“. Motivací k úpravě bylo jednak zjednodušení instrukce, ale také možnost formulovat zadání jednotlivých položek v nominativu (tedy např. Mikrofon? namísto Mikrofonem?, Stradivárky? namísto Stradivárkami? apod.). Šíře významu některých položek je však v češtině odlišná a jde buď o rozdíl kulturní (např. u položky jury – komise jsou v originálu všechny uvedené správné odpovědi vztaheny k soudnictví, což je sice v češtině označeno stejně, ale v tomto kontextu se v českém prostředí komise nevyskytuje), nebo jazykový, tedy zejména případ víceznačnosti některých slov (anglický výraz *flotsam* nemá v češtině jednoslovný ustálený název, proto se nabízí jeho volnější překlad jako trosky, což je ale v češtině mimo výrazu pro označení pozůstatků zničeného objektu také toponymum; *jib* – kosatka označuje v češtině kromě plachty také vodního živočicha), k čemuž bylo přihlédnuto při úpravě správných odpovědí. Jelikož jde o test měřící primárně znalosti, nikoliv jazykovou vybavenost probanda, mezi správné odpovědi byly dopsány i hovorové či expresivní výrazy (např. kromě na domě také na baráčku), aby bylo pro examinátory srozumitelné, že i tyto varianty vyjádření jsou přijatelné. Stejně jako v Testu 1, také v Testu 8 na základě doslovného zápisu odpovědi z pilotní a standardizační studie byly poté doplněny frekventované správné odpovědi, které byly relevantní, případně byly zařazeny k dodatečnému dotazu.

Test 12: Opakování pseudoslov

Test 12 si při adaptaci žádal hlubší úpravu. Položky byly vytvářeny v souladu s principy testu v originální verzi. Při tvorbě české verze se autoři snažili dodržet podstatu pseudoslov, tedy slovních výrazů, které svoji fonologickou a morfologickou strukturou připomínají existující slova ve slovní zásobě daného jazyka, ale nejde o ustálené kombinace hlásek (jak je tomu u slov), a tedy ze sémantického hlediska nenesou žádný význam. Při vytváření českých pseudoslov autoři dodržovali způsoby slovotvorby typické pro češtinu – zejména derivace (odvozování) a kompozice (skládání). K reálným či smyšleným slovotvorným základům byly následně přiřazeny afixy v češtině běžně používané pro derivaci substantiv, adjektiv či sloves (jejich výčet je k dispozici např. v Příruční mluvnici české (srov. Nekula, Rusínová, Grepl, & Karlík, 2012), a to jak v existujících, tak neexistujících kombinacích. Použité slabiky byly voleny tak, aby šlo o v češtině existující skupiny hlásek (str vs. *dre), což bylo ověřováno na datech v ČNK SYN2015.

Požadovaná výslovnost je uvedena u každé položky v hranatých závorkách. Přepis částečně vychází z české fonetické transkripční soustavy (srov. Palková, 2014), ale pro potřeby testu byl mírně upraven tak, aby byl srozumitelný i pro nejazykovědce. Reflektuje určitá specifika výslovnosti, včetně obměny hlásek z důvodu asimilace znělosti. Klíčové je v rámci posouzení správné kvality a kvantity hlásek rozhodně rozlišovat krátké a dlouhé samohlásky ([na] vs. [ná]) a měkkost souhlásek ([ň] znamená, že první hláska je vyslovena měkce, tedy [ň], zatímco zápis [ní] znamená, že první hláska není změkčena, tedy [n]). U spoluvýskytu dvou samohlásek (zejm. o a u) upozorňuje značka obloučku pod u (u) na to, že spojení má být vysloveno jako diftong (např. jako v existujícím českém slově [pouť] na rozdíl od [poučít]). Čárka pod r (r) pouze upozorňuje, že hláska r je v této pozici slabikotvorná.

Test 13: Audio-vizuální učení

Vznik české verze Testu 13 provázela nutnost četných úprav. Anglická slovotvorba a syntax má pravidla odlišná od češtiny. Vzhledem k obsahu položek v Testu 13 byly hlavními rozdíly, s nimiž bylo nutné se vypořádat, nezbytnost deklinace a konjugace pomocí sufixů-koncovek, absence speciální formy pro vyjádření přítomného času průběhového, tvorba negace pomocí negačního prefixu, nejednotný způsob tvorby plurálu u jednotlivých slov a fakultativnost členů v češtině. Autoři české verze se inspirovali španělskou verzí tohoto testu *Aprendizaje visual-auditivo* z Bateria III Woodcock-Muñoz (Muñoz-Sandoval, Woodcock, McGrew, & Mather, 2007), v níž byl uplatněn tvořivý přístup zejména při práci se slovesy v různém tvaru.

Deklinace substantiv, adjektiv a zájmen si v české verzi vyžádala úpravu instrukce – autoři se rozhodli ve větách použít tvary slov pro různé pády, které jsou v souladu s pravidly českého jazyka, avšak zároveň se přiklonili k toleranci v případech, používá-li proband (zejména malé děti) při čtení věty nevysložené tvary. Odpoví-li proband základním tvarem slova, je vhodné jej napoprvé ihned

opravit na vyskloňovaný či vyčasovaný tvar, aby získal zpětnou vazbu, že flexe je žádána. Nereaguje-li proband na tuto opravu, není nutné jej již znovu opravovat a je možné považovat jeho odpovědi v základním tvaru, jsou-li adekvátní, za správné.

Pro co nejmenší nutnost úpravy i kognitivní náročnost se volilo takové sloveso, jehož tvar je pro 3. osobu stejný v singuláru i plurálu.

Absence formy pro vyjádření přítomného času průběhového byla nahrazena jiným prvkem vázaným na sloveso, a to sice krátkým reflexivem. Výhodou těchto výrazů je absence gramatických rysů, z důvodu čehož zůstává jejich forma stejná pro různé osoby, čísla i rody.

Tvorba záporu si nakonec vyžádala pouze grafickou úpravu, kdy výměna pořadí a přiblížení obou symbolů naznačuje, že jde o jedno slovo sestávající ze dvou „částí“. Originální verze používá sblížení dvou symbolů v obtížnějších položkách, např. při tvorbě plurálu nebo přítomného času průběhového.

Namísto symbolu pro plurál se autoři české verze rozhodli pracovat se symbolem pro minulý čas, jehož způsob tvoření pomocí -l-ového participia zachovává jednotné paradigma a neobsahuje výjimky. Drobnou nevýhodou zvoleného symbolu (symbolů) je, že je nutná mírná úprava slovesného základu a také větší kognitivní náročnost, jelikož symbol pro minulý čas vlastně předchází symbolu pro slovo, avšak při přečtení je nejprve uvedeno sloveso a až poté koncovka pro minulý čas – tuto větší obtížnost položky nicméně ošetřují české normy.

Jako českou variantu členů autoři zvolili použití ukazovacích zájmen, která nebývají v češtině běžnou součástí vět, proto mohou působit mírně krkolomně, avšak jejichž použití je regulérní a z hlediska gramatiky správné.

Pro označení osob byla zvolena česká jména.

Dále byly autory české verze dokresleny tři symboly, jejichž použití bylo zamýšleno pro položky s nejvyšší obtížností, a z toho důvodu byly zařazeny i do finální verze. Při tvorbě symbolů byl dodržen princip testu, kdy vizuální stránka symbolu obsahuje také jistou nápovědu pro slovní označení.

Test 16: Řazení názvů a čísel

Adaptace Testu 16 do češtiny proběhla bez větších obtíží. Originální položky byly přeloženy a dle stejného principu bylo autory testu vytvořeno několik nových navíc. Zvažovalo se, zda brát v potaz odlišný počet slabik, který by teoreticky mohl mít vliv na kapacitu zapamatování, nakonec však byla dána přednost shodnosti překladu s originální verzí a obtížnost ověřena Raschovým modelem na datech z pilotní a standardizační studie.

Test 18: Paměť na slova

Originální položky z Testu 18 v celkovém počtu 26 byly přeloženy a k nim bylo pro potřeby pilotní studie vytvořeno dle stejného principu 17 nových. Z důvodu, že některé položky nebyly pro češtinu vhodné (např. některá slovesa v češtině existující pouze jako reflexiva tantum), nebyly všechny přeloženy zcela doslovně, ale s mírně posunutým významem (např. ask – žádá namísto ptá se). Po vzoru originální verze, která obsahuje slovesa v jiném tvaru nežli infinitivním (např. is, done apod.), byly také v české verzi voleny výrazy v jiném nežli základním tvaru – zejména u sloves došlo k použití různých kategorií osoby, slovesného času či způsobu (bud', dávám, dělám apod.). Použití konjugace u sloves pomohlo také k tomu, aby daná řada slov netvořila z hlediska gramatické správnosti větu, a tak nedošlo k facilitaci procesu zapamatování. Jelikož jde o test, který má primárně testovat paměť, byla ověřena známost slov, a to za použití ČNK SCHOLA2010. Všechny výrazy použité v testu jsou v tomto korpusu zastoupeny.

Vývoj škály

Každá z bank testových položek WJ IV je pro potřeby konstrukce finálního, publikovaného testu a vývoje norem kalibrována na společné škále nazvané W škála. Tato kapitola obsahuje úvod to Raschova modelu měření (Teorie odpovědi na položku, Item Response Theory, IRT) (Bond a Fox,

2007; de Ayala, 2009; Raykov a Marcoulides, 2011; Smith a Smith, 2004; Wright a Stone, 1979), který je základem pro metriku W-skóru, a také popis postupů, jimiž byly položkové banky WJ IV kalibrovány.

Využití Raschova modelu

Vývoj testu WJ IV, jeho kalibrace, vyrovnávání položkových bank a škálování byly zajištěny prostřednictvím Raschova (jednoparametrového) logistického modelu (Rasch, 1960/1980). Využití logistické funkce v analýze dichotomicky skórovaných testových položek jako první představil dánský matematik Georg Rasch. Tento model obsahuje dva parametry: schopnost člověka, B , a obtížnost položky, D , a zapisuje se jako:

$$P_{ni} = \frac{e^{B_n - D_i}}{1 + e^{B_n - D_i}}, \quad (2.1)$$

kde P_{ni} představuje pravděpodobnost, že proband n správně zodpoví položku i (Rasch, 1960/1980). Ve vzorci je pak pochopitelně použita schopnost konkrétního probanda n (B_n) a obtížnost konkrétní položky i (D_i). Pomocí pouze těchto dvou informací model predikuje, co nastane, když se proband setká s položkou. Je-li jeho schopnost vyšší než obtížnost položky (tj. jestliže rozdíl $B_n - D_i$ je kladný), potom je P_{ni} vyšší než 0,5. Naopak, je-li schopnost probanda nižší než obtížnost položky (tj. rozdíl $B_n - D_i$ je záporný), potom je P_{ni} menší než 0,5. Položky, které perfektně cílí na probandovu schopnost (tj. $B_n = D_i$), vyústí v P_{ni} rovnu 0,5.

Pro polytomicky skórované položky Wright a Masters (1982) dichotomický model později rozšířili na formu parciálního kreditu (angl. *partial credit model*):

$$P_{nik} = \frac{e^{(B_n - D_{ik})}}{1 + e^{(B_n - D_{ik})}}, \quad k = 1, 2, \dots, m_i, \quad (2.2)$$

kde P_{nik} představuje pravděpodobnost, že proband n odpoví na položku i s použitím kategorie k , a D_{ik} je obtížnost přechodu na k -tou kategorii, případně bod, na kterém se skóre v kategorii k stane více pravděpodobným než skóre v kategorii $k - 1$ ¹⁸. Model parciálního kreditu je vhodné používat u položek s různými kategoriemi odpovědí, a v důsledku toho s různými obtížnostmi (tj. různými hodnotami D_{ik}), jako např. položky z testu Psané úryvky¹⁹.

V Raschově modelu se pozorované odpovědi na položku mohou použít jako odhady obtížností testových položek, a to bez ohledu na úroveň schopnosti konkrétních probandů, kterým byl test administrován. Tento fakt se nazývá kalibrace testu nezávislá na výběrovém souboru (angl. *sample-free test calibration*). Obdobně mohou být také odpovědi na položky využity k odhadu úrovně schopnosti probandů, a to bez ohledu na obtížnosti konkrétní sady položek vyskytujících se v daném testu. To znamená, že v okamžiku, kdy je získána položková banka odpovídající modelu, je možné ke konstrukci testu použít jakoukoliv sadu položek. Takto získané skóre schopnosti respondenta v testu budou poté vztaženy k měřicí škále, na niž je škálována celá položková banka. Tato vlastnost se nazývá měření nezávislé na položkách (angl. *item-free measurement*).²⁰

Pro realizaci kalibrace testu nezávislé na výběru a měření nezávislého na položkách je nutné splnit několik předpokladů. Jedním předpokladem je *unidimenzionalita* testových položek – každá z položek musí měřit stejnou latentní schopnost. Latentními jsou tyto schopnosti nazývány proto, že nejsou měřitelné přímo. Druhým předpokladem je *lokální nezávislost*, která znamená, že probandův výkon v jedné položce není ovlivněn jeho výkonem v položce jiné. Při administraci stejné skupiny položek skupině respondentů se shodnou úrovní latentní schopnosti by položky neměly být korelovány. Třetím předpokladem je to, aby položky v testu měly *stejnou diskriminační účinnost*. To znamená, že pokud jsou nakresleny křivky pravděpodobnosti správné odpovědi pro každou položku

18 Nejjednodušší Raschův model popisuje, jak obtížné je přejít od nesprávné odpovědi k odpovědi správné. Model parciálního kreditu se používá u testů s položkami, u kterých se stupňuje míra správnosti odpovědi nebo míra souhlasu s jejich formulací. Každá kategorie odpovědi pak má vlastní charakteristickou křivku a lze definovat obtížnost, s jakou se přechází např. od „spíše souhlasím“ k „zcela souhlasím“.

19 V originále *Writing samples*, test z baterie Testy školních dovedností, který do češtiny převáděn nebyl.

20 Aby tato vlastnost platila, musí se konkrétní položky zacílit na úroveň schopnosti daných respondentů. Pokud by položky byly například příliš snadné nebo obtížné, byla by přesnost výsledných odhadů úrovně schopnosti příliš nízká na to, aby nabídla smysluplné odhady. Možnost nepřesného zacílení obtížnosti položek je značně snížena u testů, jako je například WJ IV, kde jsou položky pečlivě vybrány tak, aby pokryly celý rozsah probandových schopností.

přes rozsah schopnosti měřené testem, získané „charakteristické křivky položek“ (*item characteristic curves*, ICC) budou shodné, s výjimkou horizontálního posunutí, neboli rozdílu v obtížnosti položek, na škále latentní schopnosti.

Když tvůrci testu ověřují, zda data splňují tyto předpoklady, Raschův model poskytuje indexy shody, které mohou pomoci určit, jak dobře se data shodují s modelem. Položky, které nevykazují modelem očekávané vlastnosti, jsou buď revidovány, nebo odstraněny ze sady.

Po kalibraci položek a respondentů s pomocí Raschova modelu se výsledné odhady obtížnosti položek a schopnosti respondentů umístí na shodnou *logitovou* škálu tak, aby bylo možné úroveň schopnosti každého z respondentů popsat vzhledem k obtížnosti položek, které jsou nad a pod úrovní schopnosti probanda na dané škále. Tato logitová škála je škálou se stejnými intervaly (angl. *equal-interval scale*); vzájemné vzdálenosti mezi skóry schopnosti jsou konstantní. To znamená, že rozdíl jednoho logitu v každém bodě škály představuje stejnou změnu v respondentově schopnosti nebo obtížnosti položky. Tato vlastnost stojí v pozadí mnoha interpretačních vlastností WJ IV.

Škála W

Škála W v testech WJ IV je přímou transformací Raschovy logitové škály:

$$W = 9.1024 \text{ logitu} + 500 . \quad (2.3)$$

Tuto škálu vyvinuli Woodcock a Dahl (1971) a oproti Raschově logitové škále (Woodcock, 1999) nabízí několik praktických výhod:

1. Aditivní konstanta škálu centruje na $W = 500$, čímž odstraňuje negativní hodnoty schopnosti respondentů a obtížností položek, které se objevují v Raschově logitové metrice.
2. Znaménka obtížnosti položek a skóre schopností respondentů jsou nastavena tak, aby nízké hodnoty odpovídaly buď nízké obtížnosti položek, nebo nízké úrovni schopnosti respondenta, a naopak vysoké hodnoty vypovídaly buď o vysoké obtížnosti položek, či vysoké úrovni schopnosti respondenta.
3. Multiplikativní škálování konstantou 9,1024 vede k pravděpodobnostním důsledkům, které jsou snáze zapamatovatelné a použitelné než vzdálenosti na Raschově logitové škále.

Úroveň schopnosti respondenta a obtížnost položek jsou ve WJ IV pomocí Raschovy kalibrace položek a škálování umísťovány na společnou W škálu; stejná sada čísel může být potom použita na popis schopnosti respondenta i obtížnosti položek. Podobně jako v Raschově logitové škále má vzdálenost mezi umístěním probandovy schopnosti a umístěním obtížnosti položky na W škále (W_{A-D}) dopady na pravděpodobnost probandova úspěchu v kterékoliv položce:

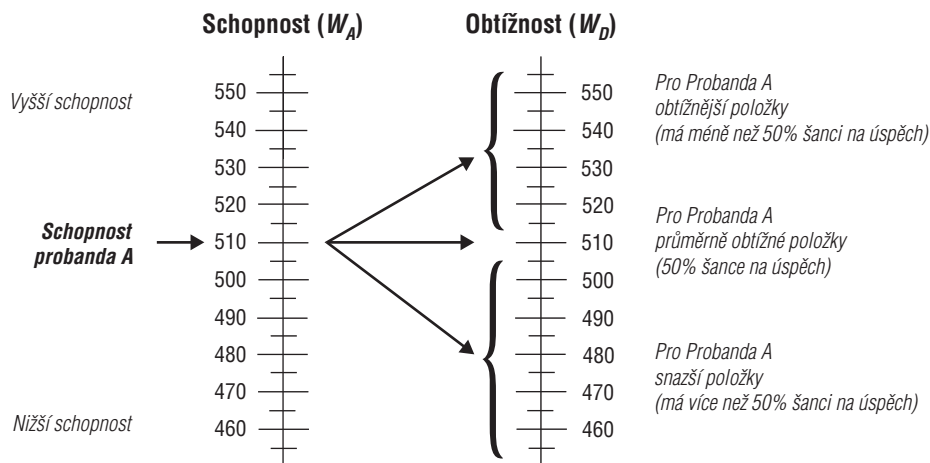
$$P_{A-D} = \frac{e^{W_{A-D}/9.1024}}{1 + e^{W_{A-D}/9.1024}} , \quad (2.4)$$

kde P_{A-D} představuje pravděpodobnost správné odpovědi asociované s danou hodnotou W_{A-D} . Obr. 2-1 ilustruje vztah mezi schopností respondenta a obtížností položky na W škále.

Na tomto grafu má proband A odhad úrovně schopnosti $W = 510$. Tento proband bude úspěšný zhruba v 50 % případů u položek s obtížností $W = 510$. Při administraci položek méně obtížných než $W = 510$ na W škále (tj. při kladné hodnotě W_{A-D}) má tento proband pravděpodobnost úspěchu v položkách větší než 50%. Obdobně má také při administraci položek obtížnějších než $W = 510$ na W škále (tj. při záporné hodnotě W_{A-D}) má proband pravděpodobnost úspěchu v položkách nižší než 50 %.

Obrázek 2-1.

Pravděpodobnost úspěchu
pro několik klíčových
hodnot $W_A - D$



Tabulka 2-1.

Pravděpodobnost úspěchu
pro několik klíčových
hodnot $W_A - D$

Schopnost probanda mínus obtížnost úkolu (W_{A-D})	Pravděpodobnost úspěchu (P_{A-D})	Schopnost probanda mínus obtížnost úkolu (W_{A-D})	Pravděpodobnost úspěchu (P_{A-D})
+50	0,996	-50	0,004
+45	0,993	-45	0,007
+40	0,988	-40	0,012
+35	0,979	-35	0,021
+30	0,964	-30	0,036
+25	0,940	-25	0,060
+20	0,900	-20	0,100
+15	0,839	-15	0,161
+10	0,750	-10	0,250
+5	0,634	-5	0,366
0	0,500	0	0,500

Škála W je podobně jako Raschova logitová škála metrikou se stejnými intervaly; vzdálenost mezi dvěma body na škále W lze interpretovat stejně na jakémkoliv úrovni schopnosti respondenta měřené testy ve WJ IV. To platí bez ohledu na to, jestli rozdíl představuje změnu ve schopnosti rozpoznat sadu písmen u předškolního dítěte či o vysokoškolskou schopnost řešit sadu příkladů z diferenciálního počtu. V obou případech platí, že je-li administrována položka, která je například o 10 bodů méně obtížná než schopnost respondenta (tj. platí, že $W_{A-D} = +10$), bude mít proband 75% pravděpodobnost úspěchu.

Pravděpodobnostní vztah, který platí pro W_{A-D} , se rovněž vztahuje i na jiné situace, ve kterých je možné určit rozdílový skór. Jde například o ukazatel úrovně schopnosti respondenta (W_A), který může být pro získání rozdílového skóru (W DIFF) porovnán s referenčním bodem na W škále (REF W). Ve WJ IV byly REF W body pro každý věk a ročník v každém testu a klastru identifikovány v procesu normalizace škály (viz kapitola 3). Standardní skóry a percentilová pořadí uvedená pro testy ve WJ IV a klustry jsou odvozeny od těchto bodů W DIFF. Další interpretační možnosti odvozené ze škály W ve WJ IV zahrnují index relativní úspěšnosti (*relative proficiency index*, RPI), instruktážní zóny a vývojové zóny, a jsou podrobněji rozebrány v Příručce administrátora WJ IV.

Vývoj položkových bank

Vývoj položkových bank pro WJ IV navázal na více než 30 let výzkumu a data z vývoje předchozích vydání testů Woodcock-Johnson²¹. Pro většinu testů existují rozsáhlé položkové banky s daty od mnoha tisíc respondentů. Pomocí Raschovy kalibrace a vyvažovacích postupů byla tato data statisticky převážena tak, aby umožnila kalibraci nových položek na existující škálu W každého z testů.

21 Jde o vývoj v USA. Data pro předchozí verzi WJ IE jsme k dispozici neměli.

V této kapitole jsou popsány postupy použité při kalibraci položkových bank, při hodnocení invariance této kalibrace a při vyvažování nových položek na existující škály W.

Ověřovací studie²²

Cíli této studie bylo: (a) sestavit položkové banky pro nové testy, z nichž bude možné vybrat formy použité v normativní studii, (b) rozšířit existující banky pro některé z testů a (c) odhadnout psychometrické charakteristiky nových položek. V případě testů, které jsou ve WJ IV nově, tato studie administrátorům nabídla také příležitost poskytnout zpětnou vazbu týkající se jasnosti testových instrukcí, materiálů, formátů odpovídání a postupů skórování.

Ověřovací formy byly sestaveny pro každý z nových testů a pro testy z WJ III, pro které byly pro potřeby WJ IV vyvinuty nové položky. Ve druhém případě testové formy z ověřovací studie obsahovaly nejen nově vyvinuté položky, ale také sadu pečlivě vybraných propojujících položek z již existujících položkových bank. Tento design propojování položek nabídl data potřebná k tomu, aby na závěr studie mohly být odhady obtížností nových položek ověřeny ve vztahu k již existujícím položkovým bankám. Pořadí obtížností položek pro každý test ve studii bylo nejprve předem odhadnuto autory testu za pomoci dat z pilotáží a zhodnocení obsahu položek či podnětového materiálu.

Ověřovací formy byly administrovány příležitostně sebraným vzorkům o velikosti od 100 do 500 probandů na každý test. Ačkoliv na to, aby tyto vzorky reprezentovaly populaci Spojených států amerických, nebyl kladen důraz, pokoušeli jsme se zajistit respondenty všech věků i schopností. Pravidla pro zahajování či ukončení testování, pokud byla použitelná, byla nastavena volněji než v typickém testování WJ IV, aby byla získána data pro více položek. Tímto způsobem bylo zajištěno získání adekvátních dat od náležitě schopných probandů i tehdy, kdy byla u položek předpokládána nesprávná úroveň obtížnosti. Po dokončení ověřovací studie byla data analyzována pomocí raschovských analytických postupů. Nové položky, které nenaplnily minimální kritéria kladená na jejich psychometrické charakteristiky, jako například korelace položky s celkovým skórem a Raschovy ukazatele fitu položky, byly poté vyloučeny z následně tvořených položkových bank.

Kalibrace položkových bank²³

Pro každý z nečasovaných testů ve WJ IV byly položkové množiny raschovsky kalibrovány a škálovány na metriku W-skóru. Většina testů WJ IV obsahuje položky, které se skórují buď jako správné (1), nebo nesprávné (0); banky položek proto byly kalibrovány pomocí dichotomického Raschova modelu. Pro testy z WJ III zahrnuté v ověřovací studii pro WJ IV byly pro odhad obtížností nových položek vzhledem k položkám z existujících bank využity postupy Raschova vyvažování pomocí společných položek (angl. *common-item equating*). Obtížnosti položek vyjádřené na škále W z kalibrovaných položkových bank poskytly informace k výběru položek do normativních verzí testů. Protože testy z WJ IV, které jsou administrovány v časovém limitu, v této fázi nebyly kalibrovány, nebyly sestaveny z položkových bank. Normativní verze časovaných testů namísto toho sestávaly z kompletní sady existujících položek pro daný test. Postupy pro kalibraci časovaných testů ve WJ IV jsou popsány níže v této kapitole.

Sestavování testů pro tvorbu norem

Výběr položek do testů WJ IV určených pro tvorbu norem byl založen na několika pravidlech. Tato pravidla jsou založena na těch, která byla použita v průběhu vývoje předchozích vydání testu, a berou také v potaz cíle revize WJ IV popsané v kapitole 1 stejně jako empirická data sebraná v průběhu fází vývoje a zkoušení položek. Některá z pravidel konstrukce byla pro většinu testů obecná, jako následující tři:

1. Položky byly vybírány tak, aby v každém testu pokrývaly rozsah obtížnosti vyplývající z požadavků na testovou baterii. U většiny testů bylo nutné, aby verze pro tvorbu norem obsahovaly na dolním konci škály položky s průměrnou obtížností pro 2- až 3leté děti a na horním konci škály položky dostatečně náročné pro bystré dospělé respondenty.

²² Tato pasáž popisuje americkou studii. Postup v případě české adaptace byl analogický.

²³ Z následující části jsou vypuštěny pasáže věnované jiným testům než testům z baterie Testy kognitivních schopností.

2. Položky byly vybírány tak, aby byly přibližně rovnoměrně rozloženy na škále obtížnosti W, s přibližně 2 a 4 body W mezi každou dvojicí položek. Hodnoty obtížnosti položek byly použity jako kritérium při výběru a uspořádání položek v každém testu.
3. Položky byly vybírány tak, aby obsah jedné položky nemohl probandovi napovědět správnou odpověď na jinou položku, ať už stejného testu nebo jiného testu z baterie WJ IV.

Sběr dat pro tvorbu norem²⁴

Normativní studie k WJ IV byla provedena v 21měsíčním období mezi červnem roku 2017 a únorem roku 2019.²⁵ Normativní data byla po převzetí od examinátorů ručně přepsána. V průběhu přepisu a před samotnými analýzami byl zkontrolován rozsah hodnot a v případě pochybností o správnosti přepisu byla data kontrolována v záznamovém archu. Byly vypočteny celkové skóry jednotlivých testů.

Položková a testová data byla během sběru normativních dat analyzována ve dvou fázích. Po sesbírání přibližně 120 případů byla nejprve provedena raschovská analýza pro ověření psychometrických charakteristik položek několika testů, které byly v češtině výrazně upravovány (viz část Vývoj české verze).²⁶ Výsledky poskytly informace, na základě kterých byly vypuštěny nevyhovující položky (zpravidla šlo o položky, které měly obtížnost velmi podobnou jiným položkám v testu, aby zbytečně nezatěžovaly probandy, ale i položky s nevyhovujícím fitem vzhledem k Raschovu modelu) a položky byly seřazeny podle vzrůstající obtížnosti.

Kalibrace a vyvažování dat z normativní studie

V případě testů s položkami skórovanými jako „správně“ či „nesprávně“ byla po dokončení studie data analyzována pomocí Raschova dichotomického modelu. K analýze dat z testů obsahujících vícebodové skórování byl použit Raschův model parciálního kreditu (angl. *partial credit model*). V případě všech testů byla data z normativní studie volně kalibrována a byly odhadnuty obtížnosti položek na škále W. Po prověření Raschových indexů fitu pro osoby (angl. *person fit indices*) byly odhaleny extrémně modelu neodpovídající odpovědi²⁷. Ty byly poté vyřazeny z kalibrace položek. Extrémně nečekané či „nefitující“ odpovědi přispívají k celkovému šumu v datech a mohou snižovat kvalitu kalibrací položek (Linacre, 2002). V tomto kroku byla u většiny testů odstraněna výrazně méně než jedna desetina procenta ze všech odpovědí²⁸.

Pro testy s existujícími položkovými bankami byly pro umístění položek na W škálu podkládající tyto položkové banky použity postupy vyvažování pomocí společných položek (angl. *common-item equating*) (Kolen a Brennan, 2010; Linacre, 2012; Wolfe, 2004; Wright a Stone, 1979). Vyvažování pomocí společných položek proběhlo s využitím následujících kroků:

1. *Identifikace stabilních kotevních položek. Standardy pro pedagogické a psychologické testování* (AERA et al., 1999) vybízí autory testů k tomu, aby prováděli „pravidelné ověřování stability stupnice, na které jsou skóry uváděny“ (s. 63). Změna obtížnosti položky v čase či napříč administracemi se nazývá posun parametru položky (angl. *item parameter drift*). Studium raschovské literatury ukazuje, že posun parametru položky mezi 0,3 až 0,5 logity má v simulovaných studiích jen malý dopad na měření probandů (Wright a Douglas, 1975; Wright a Stone, 1979), a to obzvláště tehdy, kdy je posun jednosměrný (Babcock a Albano, 2012; Stahl a Muckle, 2007). Vzhledem k tomu, že formy použité v normativní studii WJ IV obsahovaly společné kotevní položky vybrané z rozsáhlých před-kalibrovaných položkových bank v kombinaci s položkami nově vyvinutými, bylo možné porovnat obtížnosti na W škále pro tyto propojující položky s obtížnostmi z předchozí kalibrace položkové banky. Tím mohla být ověřena stabilita odhadů obtížnosti položek v čase. Pro tyto společné položky byly

²⁴ Tato část je věnována sběru dat české adaptace WJ IV COG.

²⁵ Tato sekce obsahuje informace o sběru dat na úrovni položek. Kapitola 3 obsahuje podrobné informace o tréninku examinátorů, standardizačních postupech a metodách sběru dat, které byly použity v normativní studii, jakož i informace o vývoji norem k WJ IV.

²⁶ Provedeny byly také předběžné strukturní analýzy baterie. Ty jsou popsány v kapitole 5.

²⁷ Příkladem takové odpovědi je chybná odpověď na první nebo druhou položku testu respondentem, který jinak v testu dosáhl vysokého skóru. Taková odpověď je patrně způsobena počátečním nepochopením instrukce nebo nepozorností a nesouvisí s měřeným latentním rysem.

²⁸ Platí i pro česká data.

volně odhadnuté obtížnosti položek na škále W z normativního datového souboru zakresleny do grafu oproti obtížnostem kalibrovaných položek z předcházejících položkových bank. Tyto grafy odhalily, které položky (pokud vůbec některé) měly v obou datových souborech zřetelně odlišné relativní polohy obtížnosti (tj. pořadí). Pro účely této analýzy byl za zřetelný posun parametru položky považován takový rozdíl mezi kalibrací položkové banky a normativní kalibrací WJ IV, který na škále kalibrace položkové banky činil zhruba 5 W bodů (o něco více než 0,5 logitu) a více. Tyto položky byly pro účely vyvažování vyloučeny z možnosti figurovat jako spojovací položky a byly v následných kalibracích považovány za položky nové. Ve většině testů byly takto identifikované položky přibližně rovnoměrně rozloženy z hlediska pozitivního a negativního posunu.

2. *Provedení kotvené (angl. anchored) kalibrace na normativních datech.* Po identifikaci stabilní sady kotevních položek byly obtížnosti těchto položek na škále W v analýze programu WINSTEPS® (Linacre, 2012) zahrnující všechny normativní položky „ukotveny“ k hodnotám obtížnosti své položkové banky. Tímto způsobem byly získány hodnoty obtížnosti položek na škále W, na které je založena položková banka, pro všechny normativní položky včetně těch nově vyvinutých.

Kalibrace časovaných testů

Pro každý z časovaných testů ve WJ IV²⁹ byla použita nová kalibrační procedura založená na Raschově modelu posuzovací škály (angl. *rating scale model*) (Andrich, 1978). Postup sestával z následujících dvou kroků:

1. *Převedení hrubých skóre na metriku založenou na rychlosti* (angl. *rate-based metric*). Tímto způsobem byl převáděn probandův hrubý skóre za minutu zaznamenaný administrátory na poměrné skóre pro každou minutu testování.
2. *Získání Raschových odhadů obtížnosti pro každou minutu testování.* Tyto poměrné skóre jsou následně analyzovány Raschovým modelem posuzovací škály. Tento model se zapisuje jako:

$$P_{nik} = \frac{e^{(B_n - (D_i + F_k))}}{1 + e^{(B_n - (D_i + F_k))}}, k = 1, 2, \dots, m_i, \quad (2.5)$$

kde P_{nik} představuje pravděpodobnost, že proband n obdrží skóre v kategorii k položky i , D_i je obtížnost i -té položky a F_k umístění prahu kategorie, neboli bod, ve kterém je pravděpodobnější, že proband obdrží spíše skóre v kategorii k než v kategorii $k - 1$ ³⁰. V této aplikaci je analýzou každá minuta testování považována za položku; prahy kategorií jsou pozice na škále W, kde se rychlostní skóre k stává více pravděpodobným než rychlostní skóre $k - 1$. Proband, jemuž je administrován test Fluence psaní vět³¹ (angl. *Sentence Writing Fluency*), tak například obdrží skóre ze sedmi různých „položek“ či minut, a každý z těchto skóre odráží jeho rychlost správného odpovídání v konkrétní minutě testování.

Kalibrace založená na rychlosti v časovaných testech WJ IV přirozeně odměňuje ty, kdo pracují rychle a přesně; eliminuje navíc potřebu pracovat s „bonusovými body“, které byly skórovacími algoritmy ve WJ III přidělovány těm respondentům, kteří odpovídání ukončili brzy. Tato metoda navíc přirozeně penalizuje ty probandy, kteří odpovídají rychle, ale nepozorně. Výhody tohoto Raschova modelu pro práci s rychlostními daty převažují nevýhodou spočívající v tom, že data z těchto rychlostně skórovaných testů WJ IV nesplňují předpoklad lokální nezávislosti Raschova modelu (McGrew, Werder a Woodcock, 1991). Zaprvé, tato metoda poskytuje odhady obtížnosti a úrovně schopnosti respondentů na škále W, které mohou být posléze kombinovány s jinými W-skóre z jiných testů z WJ IV pro vytvoření klastrových kompozitních skóre. Zadruhé, metoda rozlišuje mezi respondenty pracujícími přesně a rychle a těmi, kdo pracují přesně, ale méně rychle, tím, že ty první odmění vyššími W-skóre. A konečně časované testy byly kalibrovány na datech sebraných na základě

29 Výraz „časované testy“ odkazuje na testy či subtesty, které po probandech žádají odpovídání (písemné či ústní) při stanoveném časovém limitu. V baterii WJ IV COG jde o testy/subtesty Hledání písmen, Fonologické zpracování – Slovní fluence, Hledání čísel a Vyhledávání dvojic.

30 U předchozího typu modelu je poznámka pod čarou, která model vysvětluje blíže. U tohoto modelu je vysvětlení analogické.

31 Tento test není součástí baterie Testů kognitivních schopností, ale baterie Testů školních dovedností, která nebyla do češtiny převezena.

administrace forem WJ IV v normativní studii; neexistují žádné položkové banky, a v důsledku toho je ani není nutné vyvažovat pro získání společné škály. Z toho plyne, že tyto testy nevyžadují naplnění vlastnosti měření nezávislého na testu Raschova modelu. Model byl namísto toho v této aplikaci použit k získání odhadů úrovně schopností respondentů v normativní studii, aby je poté bylo možné použít při konstrukci norem WJ IV a tvorbě převodních tabulek mezi hrubými skóry a W-skóry obsažených v publikovaných formách časovaných testů.³²

Analýzy zkreslení položek³³

Zkreslení testu odkazuje k situaci, kdy nedostatky „v testu samotném nebo ve způsobu, jakým je užíván, vedou k rozdílnému významu skóre získaných příslušníky různých kategorií osob“ (AERA et al., 1999, 2001, s. 80). Zkreslení v obtížnosti položky, často označované jako diferenciální fungování položky (angl. *differential item functioning*, DIF), vzniká, když je položka obtížnější pro určitou podskupinu respondentů, přestože celková úroveň schopnosti těchto respondentů je stejná jako u jiných skupin. Pro položky WJ IV použité při tvorbě norem bylo DIF analyzováno v průběhu kalibrace položek za pomoci Raschovy iterativní logitové metody (angl. *iterative-logit method*) v software WINSTEPS (Linacre, 2012). V této metodě jsou kalibrace obtížnosti položek a jejich příslušné standardní chyby odhadovány pro každou položku a každou podskupinu zvlášť, zatímco odhady obtížnosti všech ostatních položek jsou drženy jako konstantní. Rozdíl mezi odhady obtížnosti položky v podskupinách pro každou položku, neboli *DIF kontrast*, lze hodnotit s použitím Welchovy t-statistiky pro rozdíl dvou průměrů (Linacre, 2012).

Všechny položky použité při tvorbě norem všech testů WJ IV, kromě časovaných testů, byly hodnoceny z hlediska DIF na základě pohlaví. V americké verzi WJ IV bylo DIF položek hodnoceno i pro rasu a etnicitu, což v případě homogenního českého výběrového souboru není možné. Položky vykazovaly DIF, pokud byla absolutní hodnota DIF kontrastu větší než 0,5 logitů (tzv. věcná významnost) a současně statisticky významný výsledek Welchova t-testu (jejich četnosti jsou uvedeny v tab. 2-2). Četnosti položek se statisticky i věcně významnými výsledky jsou poměrně nízké, srovnatelné s výsledky statistických analýz amerického standardizačního souboru. V českém souboru je nejvíce takových položek v subtestu T8: Všeobecné znalosti – K čemu, což je výsledek podobný jako v americkém technickém manuálu, kde autoři konstatují, že vyšší podíl DIF se vyskytuje v testech syčených *Gc* než v testech syčených *Gf*. Čtyři položky byly snadnější pro muže než pro ženy se stejnou úrovní všeobecných znalostí – potrubí, tachometr, šuplera a síkovky, naopak u čtyř položek byly správné odpovědi častější u žen – díář, slovník, stradivářky, kardamom. Domníváme se, že tyto rozdíly lze poměrně snadno interpretovat obecně větším zaměřením mužů na technické záležitosti. Vzhledem k tomu, že čtyři položky vykazují DIF ve směru větší obtížnosti u mužů a čtyři u žen, tento test podle našeho názoru celkově nevykazuje zkreslení.

32 Ačkoli jsou časované testy kalibrovány na W škálu schopnosti z praktických důvodů, uživatelé testu by si měli být vědomi několika faktorů ovlivňujících interpretační komplexitu rozdílových W-skóre a indexů relativní úspěšnosti (RPI) pro tyto testy. RPI je skóre odkazující ke kritériu (angl. *criterion-referenced*), který poskytuje funkční informace o probandově výkonu. Zatímco normativní skóre (angl. *norm-referenced*) jako standardní skóre nebo percentilová pořadí popisují relativní umístění probanda v rámci skupiny, RPI popisuje kvalitu jeho výkonu v posuzovaných dovednostech či schopnostech ve srovnání s výkonem osob z normativního vzorku srovnatelných s probandem co do věku či vzdělání. RPI je určen rozdílovým W-skórem, neboli vzdáleností probandova skóre od referenčního bodu na W škále pro daný věk či vzdělání (pro podrobnější vysvětlení RPI a jeho využití viz Jaffe [2009]).

Vzhledem k tomu, že RPI a W-skóre poskytují jinou úroveň informace, neočekává se, že by mezi rozdílovým skórem W a RPI a normativními skóry jako standardními skóry a percentilovým pořadím byl lineární vztah. Rozdílové W skóre a RPI odpovídající danému standardnímu skóru nebo percentilovému pořadí se dokonce mohou napříč věky a ročníky školní docházky znatelně lišit kvůli rozložení latentní měřené schopnosti v konkrétním věku či ročníku. Tyto rozdíly skóre odrážejí rozdílnou míru růstu latentního rysu zkoumaného daným testem. Obecně vzato vykazují míry úrovně schopnosti probandů získané metodami založenými na rychlosti v časovaných testech WJ IV větší směrodatné odchylky napříč věky, než tomu bylo u časovaných testů WJ III. Časované testy ve WJ IV mají navíc napříč věky také větší směrodatné odchylky než testy nečasované, což je způsobeno tím, že kalibrační metoda založená na rychlosti vytváří škály W rozdávající participanty do mnohem více úrovní schopnosti. Z toho plyne, že vzhledem k senzitivitě rozdílových W-skóre a RPI vůči míře růstu latentního rysu může být mezi SS/PR skóry (standardními a percentilovými) a RPI a W-skóry u časovaných testů větší míra divergence.

33 Úvodní odstavec je překladem z originálního amerického manuálu (přeložil Tomáš Urbánek), druhý odstavec je upraven vzhledem ke specifikům české adaptace WJ IV COG a tabulka 2-2 je založena na analýzách českých dat.

Tabulka 2-2.
*Procenta položek
s potenciálním DIF*

Název testu / subtestu	Počet položek	Počet položek obtížnějších pro		Podíl položek obtížnějších pro	
		Muže	Ženy	Muže	Ženy
T01 Slovník – Synonyma	29	3	2	10,3%	6,9%
T01 Slovník – Antonyma	30	4	2	13,3%	6,7%
T02 Číselné řady	42	0	1	0,0%	2,4%
T03 Verbální pozornost	37	1	2	2,7%	5,4%
T05 Fonologické zpracování – Vyhledávání slov	24	1	0	4,2%	0,0%
T05 Fonologické zpracování – Substitute	19	1	1	5,3%	5,3%
T06 Reprodukce příběhů	216	0	0	0,0%	0,0%
T07 Vizualizace – Prostorové vztahy	24	0	0	0,0%	0,0%
T07 Vizualizace – Rotace s kostkami	24	1	1	4,2%	4,2%
T08 Všeobecné znalosti – Kde	22	1	2	4,5%	9,1%
T08 Všeobecné znalosti – K čemu	22	4	4	18,2%	18,2%
T09 Formování konceptů	40	0	1	0,0%	2,5%
T10 Obrácené číselné řady	34	0	2	0,0%	5,9%
T12 Opakování pseudoslov	47	0	2	0,0%	4,3%
T13 Audio-vizuální učení	190	11	3	5,8%	1,6%
T14 Rozpoznávání obrázků	24	1	1	4,2%	4,2%
T15 Analýza – Syntéza	35	2	1	5,7%	2,9%
T16 Řazení názvů a čísel	31	0	1	0,0%	3,2%
T18 Paměť na slova	26	1	1	3,8%	3,8%

Standardizace a vývoj norem WJ IV

Tato kapitola obsahuje informace o normalizaci Woodcock-Johnson IV Testů kognitivních schopností a výpočtech normovaných skóre pro všechny testy a klastry. V souladu se *Standards pro pedagogické a psychologické testování* (AERA et al., 1999, 2001) přináší tato kapitola popis výběrového souboru, u kterého byly počítány normy, popis procesů sběru dat a statistické metody použité při výpočtech normovaných skóre. Tyto informace přinášejí doklady o tom, že normované skóre WJ IV COG lze použít k popisu kognitivních schopností probandů od předškolního věku po stáří.

Obecné charakteristiky normalizačního souboru

Složení výběrového souboru bylo naplánováno tak, aby v něm byly určité věkové skupiny (děti a adolescenti) nadreprezentovány (viz tab. 3-2), tzn. byl vybrán větší počet jejich zástupců, než kolik by odpovídalo poměrnému zastoupení v populaci, aby tak bylo možné přesněji odhadnout jejich statistické charakteristiky. Nejde tedy o reprezentativní výběr z české populace, ale jak je vidět z tabulek 3-1 a 3-2, jsou v něm zastoupeny všechny důležité věkové skupiny, skupiny podle vzdělání rodičů a vzdělání respondentů. Zastoupení mužů a žen v jednotlivých věkových skupinách je přibližně rovnoměrné.

Fakt, že ve standardizační studii nebyl získán reprezentativní výběr z populace ČR, není vzhledem k vlastnostem Raschova modelu nijak na překážku adekvátní kalibraci položek jednotlivých testů baterie WJ IV a tedy adekvátnímu měření příslušných schopností. V důsledku nereprezentativnosti výběru na základě norem není možné usuzovat na charakteristiky populace ČR – pouze na vlastnosti testovaných respondentů.

Nábor examinátorů byl prováděn několika formami. Oslovili jsme účastníky školení WJ IE o spolupráci při standardizaci nové verze metody, a to na základě seznamů frekventantů školení, (školiiteli byli PhDr. Anton Furman, CSc. a PhDr. Jarmila Burešová). Uveřejnili jsme inzerát v Zpravodaji ČMPS. Informace o náboru byla rovněž uveřejněna na úvodní stránce nakladatelství Propsyco.

Všichni examinátoři, kteří prováděli vyšetření metodou WJ IV COG, byli zaškoleni v administraci v jednodenním kurzu. V rámci tohoto školení byli frekventanti požádáni, aby informovali své kolegy a kolegyně o možnosti účasti na standardizaci metody.

Examinátoři se v průběhu sběru dat hlásili přes webovou aplikaci, kde si „rezervovali“ určitého respondenta (např. žena ve věku 21–25 let se SŠ vzděláním). Tím se udržoval přehled o tom, které kategorie respondentů přibývají, a kde je naopak potřeba soustředit úsilí při sběru dat.

Soubor dat, na základě kterého byly prováděny kalibrace položek v jednotlivých testech a výpočty norem, tvoří odpovědi od 936 probandů. Tvoří ho 479 (51,2 %) chlapců a mužů a 457 (48,8 %) dívek a žen. Celkový průměrný věk je 25,3 roku při směrodatné odchylce 18,95, medián věku je 18,4, minimální věk je 2 roky a maximální 80 let (tab. 3-1).

Tabulka 3-1.

Věkové charakteristiky
standardizačního souboru

	m	sd	medián	min	max
muži	24,5	18,40	18,1	2,1	79,9
ženy	26,1	19,50	18,9	2,0	80,0

Tabulka 3-2.

Vzdělání rodičů a vzdělání
respondentů

věk	N	muži	ženy	vzdělání				vzdělání rodičů			
				ZŠ nebo vyučení	Maturita	VŠ	chybí	ZŠ nebo vyučení	Maturita	VŠ	chybí
2;0 – 2;11	24	12	12	0	0	0	24	10	10	4	0
3;0 – 4;11	44	22	22	0	0	0	44	23	14	7	0
5;0 – 6;11	64	32	32	0	0	0	64	27	23	13	1
7;0 – 8;11	57	33	24	0	0	0	57	22	22	12	1
9;0 – 10;11	48	27	21	0	0	0	48	17	19	11	1
11;0 – 12;11	62	33	29	0	0	0	62	26	24	12	0
13;0 – 14;11	60	30	30	0	0	0	60	29	20	11	0
15;0 – 16;11	61	32	29	0	0	0	61	27	22	12	0
17;0 – 18;11	64	34	30	0	0	0	64	30	20	12	2
19;0 – 20;11	60	29	31	0	0	0	60	27	22	10	1
21;0 – 25;11	42	22	20	15	19	5	3	0	1	0	41
26;0 – 30;11	46	24	22	15	20	11	0	0	0	0	46
31;0 – 40;11	93	47	46	40	32	19	2	0	0	0	93
41;0 – 50;11	90	47	43	43	32	15	0	0	0	0	90
51;0 – 60;11	60	28	32	35	17	7	1	0	0	0	60
61;0 – 70;11	42	20	22	24	14	4	0	0	0	0	42
71;0 – 80;11	19	7	12	13	4	2	0	0	0	0	19
celkem	936	479	457	185	138	63	550	238	197	104	397

Postupy normalizační studie

Jak bylo uvedeno, examinátoři byli psychologové, kteří měli většinou zkušenosti s administrací a vyhodnocováním předchozí verze baterie WJ IE. Byli vyškoleni členy týmu dr. Furmanem a dr. Burešovou. V průběhu sběru dat byly monitorovány charakteristiky vznikajícího datového souboru. Průběžně byl prováděn přepis dat a kontrola případných chyb.

Tvorba norem

Data od 936 respondentů byla zpracována v řadě kalibračních analýz, kdy byly vybrány finální sady položek použité hlavně v testech, které bylo potřeba v české verzi modifikovat nebo nově vytvořit. Pro všechny testy pak byly vypočteny celkové skóre a na základě finálních kalibrací položek v programu Winsteps (Linacre, 2012) W-skóre a příslušné standardní chyby.

Tyto W-skóre pak byly použity jako vstupní data pro tvorbu norem, kdy byly v knihovně cNORM (Lenhard & Lenhard, 2019) ve statistickém software R (R Core Team, 2019) nejprve hledány optimální parametry polynomů, které by bylo možné použít pro jejich vyhlazování, a poté vygenerovány normy.

Souhrnné statistiky testů a klastřů

Souhrnné statistiky a reliability pro testové a klastrové skóry WJ IV COG jsou uvedeny pro 15 věkových podskupin v přílohách B a C. Jsou zde uvedeny průměry a směrodatné odchylky původních W-skórů, standardní chyby měření W-skórů a standardních skórů a reliability odhadnuté na základě výsledků kalibrace Raschových modelů pro jednotlivé testy a klastry.

Tvorba a výpočet klastrových skórů WJ IV

S výjimkou klastru Obecné intelektové schopnosti (GIA) jsou všechny klastrové skóry vypočteny jako aritmetický průměr W-skórů testů, kterými jsou tyto klastrové skóry tvořeny. Např. klastrový skór klastru Porozumění-znalosti (Gc) tvořeného testy Slovník (T1) a Všeobecné znalosti (T8) se počítá jako

$$\frac{\text{W skór testu Slovník} + \text{W skór testu Všeobecné znalosti}}{2}$$

Výpočet diferenciálně váženého klastrového skóru GIA

Klastrový skór Obecné intelektové schopnosti (GIA) je počítán na základě diferenciálně váženého skóru. Byla provedena analýza hlavních komponent pro získání vah g-faktoru přes celý rozsah výběrového souboru pro testy, které do skóru GIA přispívají. Tyto váhy jsou uvedeny v tabulce 3-3.

Tabulka 3-3.

Váhy pro výpočet skóru GIA
v české standardizaci

Test 1	Slovník	0,145
Test 2	Číselné řady	0,146
Test 3	Verbální pozornost	0,145
Test 4	Hledání písmen	0,144
Test 5	Fonologické zpracování	0,145
Test 6	Reprodukce příběhů	0,137
Test 7	Vizualizace	0,138

Skór Obecné intelektové schopnosti (GIA) ve WJ IV COG je skórem obecné inteligence (g); představuje první hlavní komponentu získanou z analýzy hlavních komponent (PCA). Váhy GIA určené pomocí PCA byly poprvé použity k určení skóru v původním klastru Široká kognitivní schopnost (BCA) v baterii WJ (Woodcock, 1978). V baterii WJ-R měl klastr Širokých kognitivních schopností (BCA) skór vytvořený s použitím vah, které byly získány stejným způsobem (McGrew et al., 1991). Váhy pro výpočet GIA určené pomocí PCA byly znovu zavedeny do baterie WJ III (McGrew & Woodcock, 2001). V PCA se hledá optimální vážená kombinace testů, která vysvětluje největší podíl rozptylu v sadě testů, konkrétně první taková kombinace, první komponenta (podobná faktorů ve faktorové analýze). Použití vah z PCA zajišťuje, že váhy všech jednotlivých testů jsou optimální. Oproti tomu v ostatních bateriích testů kognitivních schopností (včetně Wechslerových škál inteligence a WJ-R) přispívají všechny testy tvořící skór obecné inteligence (g) stejnou vahou. Použití PCA pro měření kognitivních schopností jako základ pro určení různých vah testů vede k nejlepšímu statistickému odhadu obecné inteligence (Jensen, 1998).

Konstrukce norem a odvozených skórů WJ IV³⁴

Jak je popsáno v *Technickém manuálu Woodcock-Johnson III* (McGrew & Woodcock, 2001) a v *Technickém manuálu Woodcock-Johnson III Aktualizace norem* (McGrew, Schrank & Woodcock, 2007), vývoj testových norem a odvozených skórů vyžaduje stanovení „normativního“ (průměrného) skóru pro každé měření v každé věkové skupině, pro kterou se plánují normativní interpretace. V rodině nástrojů WJ se takovému normativnímu skóru říká Referenční W-skór (REF W). Když jsou zobrazeny v grafu jako funkce chronologického věku, skóry REF W mají podle předpokladu charak-

34 Následující pasáže věnované tvorbě norem v americké verzi baterie WJ IV COG jsou přeloženy z původního manuálu Petrem Pališkem.

teristiky vývojových křivek (viz růstové křivky klastrů v kapitole 5). Tyto testové a klastrové křivky skóru REF W jsou vizuální grafické reprezentace průměrných výkonů účastníků standardizační studie v každém věku sloužící efektivnímu využití konkrétního měřicího nástroje.

Křivky REF W slouží jako základ pro tvorbu věkově a ročníkově ekvivalentních skóru a index relativní úspěšnosti (RPI), které rozvíjejí interpretační vlastnosti WJ IV. Jsou-li navíc směrodatné odchylky (SD) skóru v každém věku naneseny jako funkce věku nebo ročníku, představují výsledné křivky hodnoty SD, které ve spojení s REF W tvoří základ pro výpočet ostatních normativních skóru (např. standardních skóru nebo percentilových pořadí).

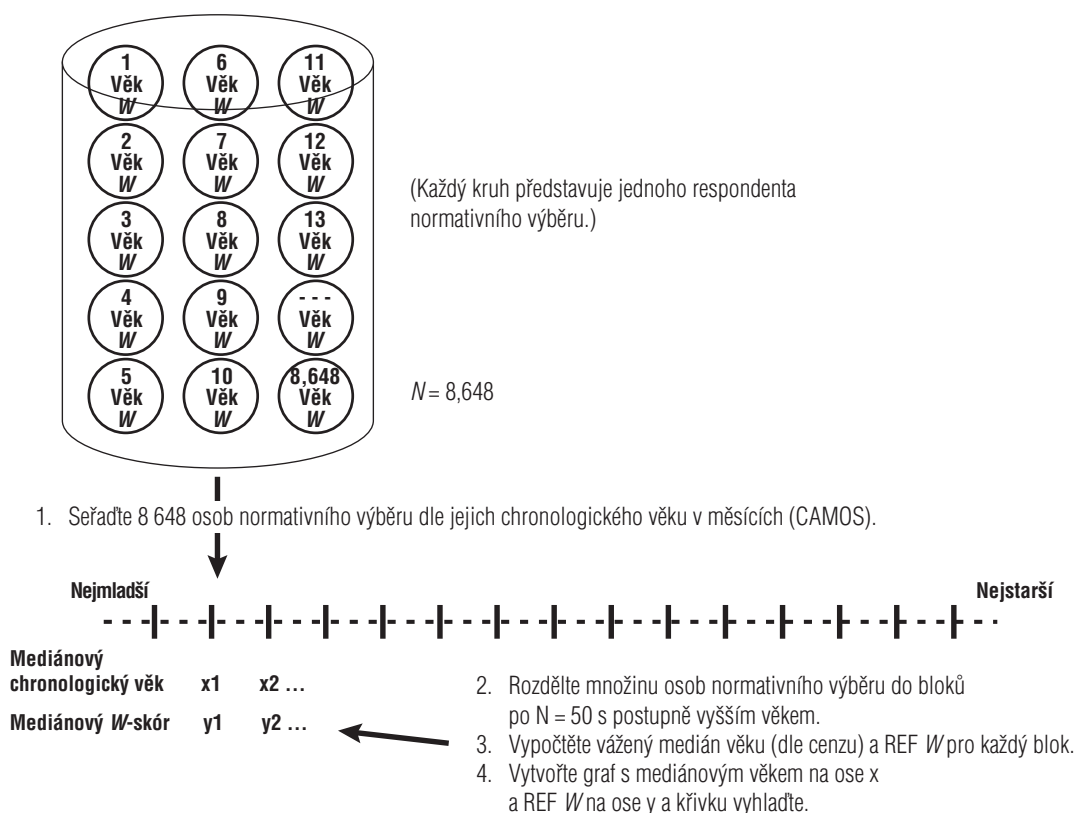
Postupy použité ke konstrukci norem ve WJ IV odpovídají těm, které byly použity u WJ III NU. Tyto inovativní procedury, které do oblasti konstrukce testových norem přivedly metodologii bootstrapaného resamplingu, jsou podrobně popsány McGrewem, Daileym a Schrankem (2007). Následující popis a grafy včetně prezentace výsledků testu Rozpoznávání písmen a slov jsou ze zmíněného dokumentu z roku 2007. Dále jsou nejprve stručně popsány tradiční metody výpočtu hodnot REF W pro nástroje z rodiny WJ (viz McGrew a Woodcock, 2001), načež následuje představení metod založených na bootstrapu, jež jsou používány v WJ III NU a WJ IV. Nakonec jsou vysvětleny výhody této novější metody³⁵.

Popis tradičního postupu výpočtu norem

Ve WJ III a dřívějších vydáních byly hodnoty REF W každého z nástrojů získávány z vyhlazených křivek procházejících datovými body vzešlými ze vzorku, jež představovaly průměrné hodnoty REF W pro seřazené skupiny či bloky po 50 účastnících normativní studie. Tradiční proces (bez bootstrapu), jenž byl využit při výpočtech pro normy ve WJ III, je zdokumentován na Obrázcích 3-1 a 3-2 na příkladu testu Rozpoznávání písmen a slov.

Obrázek 3-1.

Tradiční proces výpočtu norem WJ III ilustrovaný na příkladu testu Rozpoznávání písmen a slov – kroky 1 až 4.



Poznámka: Celková velikost výběru (8 648) v tomto obrázku představuje normativní výběrový soubor WJ III.

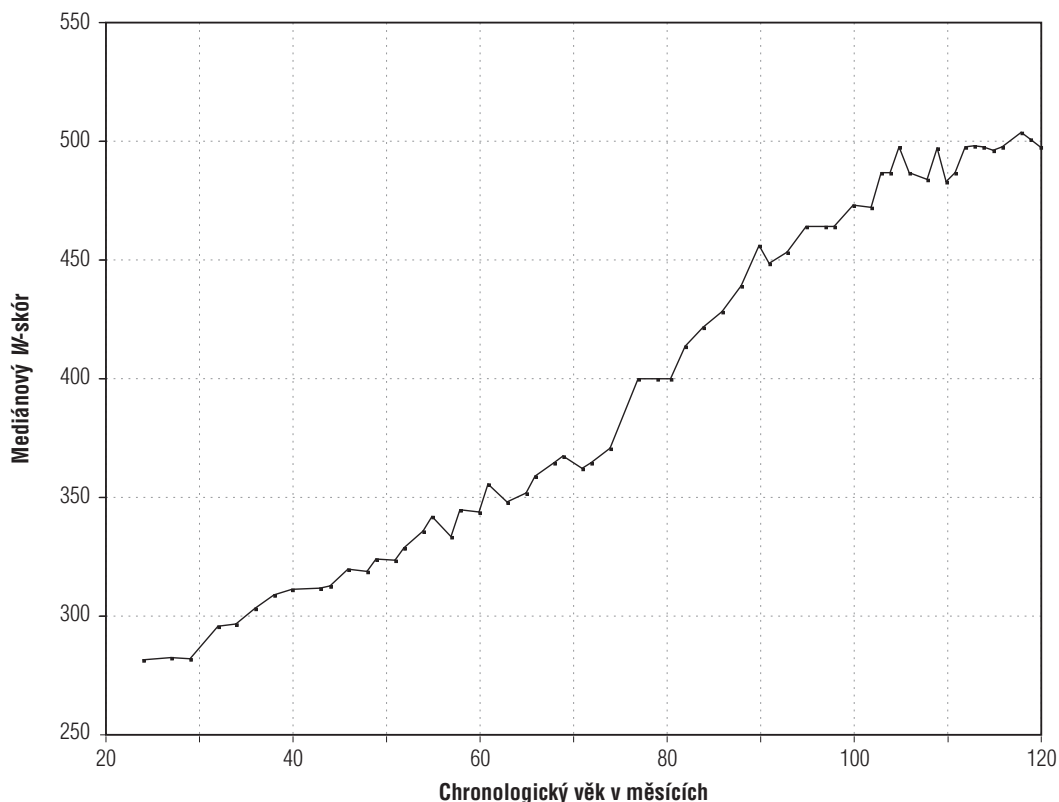
35 V americké verzi byl při tvorbě norem použit zde popsáný postup. Nástroje pro polynomické vyhlazování, které jsme použili v české verzi, nebyly v době přípravy americké verze pravděpodobně k dispozici. Tuto pasáž uvádíme z toho důvodu, že dobře ilustruje princip, na základě kterého jsou normy vytvářeny. Námí použitý sofistikovanější algoritmus využívající knihovnu cNORM směřoval k analogickým cílům.

Prvním krokem z Obrázku 3-1 je seřazení příslušných skóre všech participantů dle jejich chronologického věku, což je zde reprezentováno válcem. Přímka pak znázorňuje řazení od nejmladšího po nejstaršího (od levého po pravý konec přerušované přímky). Dalším krokem je pak postupné rozdělení seřazeného souboru účastníků do skupin či bloků o 50 osobách (v Obrázku 3-1 je každý z nich znázorněn plochou mezi sousedními svislými značkami na přímce). První soubor nejmladších účastníků je tak na levém kraji přerušované přímky představující všechny seřazené participanty. Výsledkem je soubor bloků seřazený dle věků příslušných osob od nejmladších po nejstarší.

Třetím krokem z Obrázku 3-1 je poté výpočet váženého chronologického věku (v měsících) a W skóru pro každý blok, a to za užití vah participantů určených pomocí procedur popsanych výše. Takto vzniklé páry věku a hodnot W-skóru pro každý z bloků pak představují surový materiál pro nanášení křivky REF W do grafu a výpočet jejich hodnot pro zvolený test (viz Obrázek 3-2³⁶).

Obrázek 3-2.

Graf vybraných věků (20–120 měsíců) a mediánových W-skóre jednotlivých bloků pro test Rozpoznávání písmen a slov (viz krok 3 v obr. 3-1).



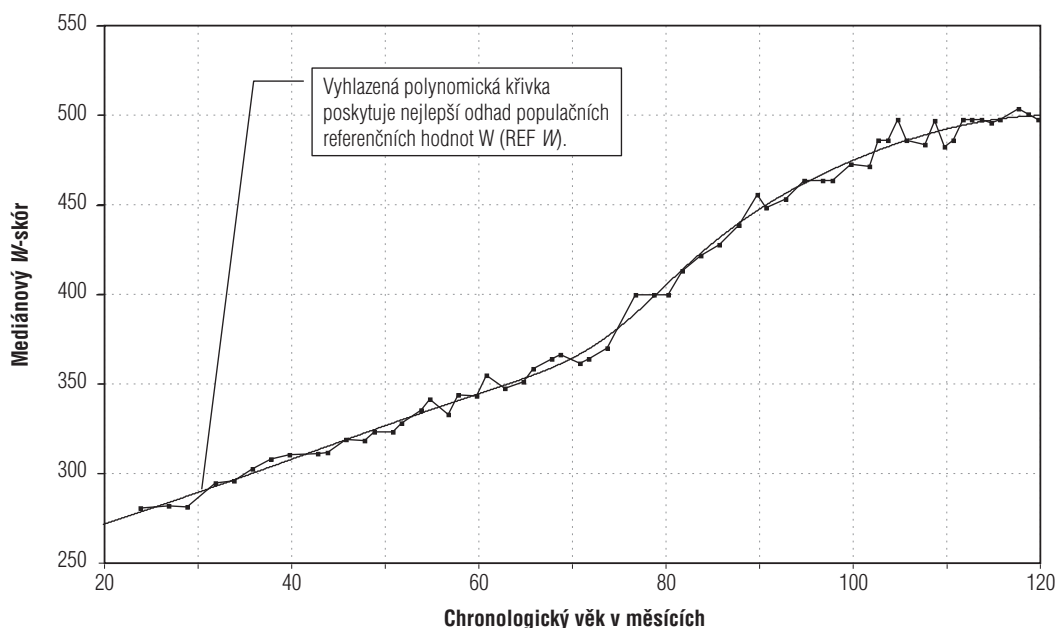
Z Obrázku 3-2 je patrné, že i když hodnoty REF W sledují konzistentní růstový trend, objevuje se také určitý „šum“, jenž je způsoben výběrovou chybou (Woodcock, 1994). K odstranění této chyby a dosažení „vyhlazeného“ řešení jsou využívány postupy, které umožňuje specializovaný software zaměřený na prokládání polynomiálních křivek a poskytování co nejlepších odhadů populačních hodnot parametru REF W (McGrew a Wrightson, 1997; Woodcock, 1994). Obrázek 3-3 představuje výsledek těchto procedur aplikovaných na data z testu Rozpoznávání písmen a slov z WJ III zobrazená v Obrázku 3-2. Vyhlazená křivka poskytuje normativní hodnoty REF W, jež byly použity k odvození skóre ve WJ III³⁷ (např. věkových a ročníkových ekvivalentů, RPI, standardních skóre a percentilových pořadí). Vykreslovací a vyhlazovací proces je poté opakován pro směrodatné odchylky založené na datech z výběru³⁸.

36 Z ilustrativních důvodů jsou datové body věku po REF W v obrázku prezentovány jen mezi věky 20 až 120 měsíců. V praxi jsou křivky REF W prokládány napříč celým věkovým rozsahem.

37 Vyhlazená křivka uvedená na tomto obrázku je pouze ilustrativní a nepředstavuje finální podobu křivky použitou při konstrukci norem WJ III.

38 Prameny nabízející vysvětlení konstrukce norem skrze prokládání křivek jsou kupříkladu Daniel (1997), Gorsuch a Zachary (1985), McGrew a Woodcock (2001), McGrew a Wrightson (1997) a Woodcock (1994).

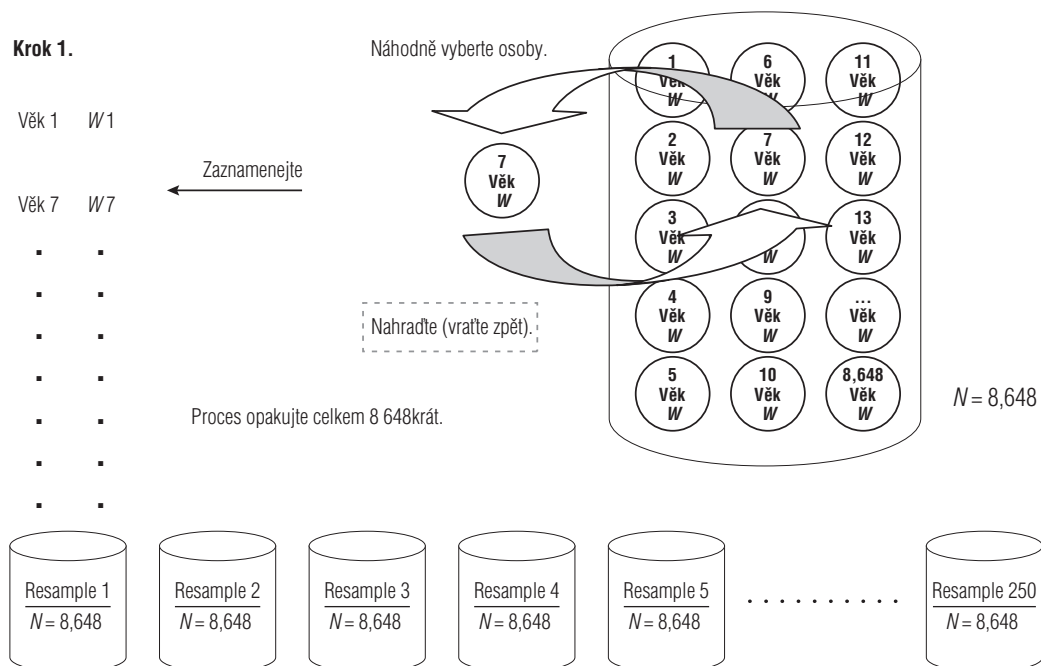
Obrázek 3-3.
Vyhlazené řešení
polynomické křivky
z Obrázku 3-2.



Poznámka: Podobný proces následně proběhne se směrodatnými odchylkami (SD) subtestu Rozpoznávání písmen a slov.

Metoda bootstrapu spočívá v tvorbě empirických rozložení nějakého statistického indexu vypočtených z výběrového souboru osob náhodně vybraných z populace. Variabilitu indexu v tomto rozložení lze interpretovat jako rozsah, do něhož by spadala skutečná hodnota tohoto indexu, pokud bychom změřili celou populaci. Zjednodušeně řečeno bootstrapování, aplikováno na výpočet datových bodů věku a W-skóru při prokládání křivkou, vytváří kolem každého bodu interval spolehlivosti nebo pásmo, obdobně jako by examinační používali výběrové chyby nebo pásma spolehlivosti k ohraničení individuálních testových skóre. Empirické rozložení daného indexu je tvořeno opakovaným vytvářením nových výběrů (*resampling*) z obdrženého normativního výběrového souboru a přepočtem požadované statistiky pro každý opakovaný výběr (*resample*).

Obrázek 3-4.
Tvorba 250 nových výběrů
pro test Rozpoznávání
písmen a slov z WJ III NU
pomocí náhodného výběru
probandů s nahrazením
(bootstrapová metoda).



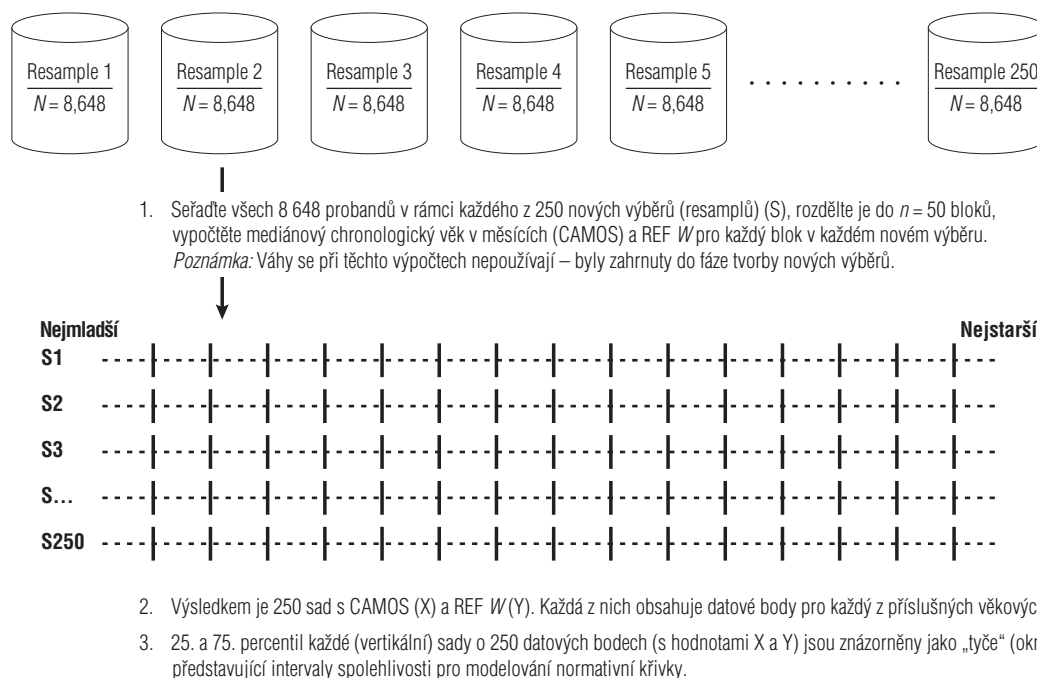
V případě postupu při výpočtu norem ve WJ IV (a WJ III NU) bylo ze souboru všech účastníků normativní studie vytvořeno 250 nových opakovaných výběrů³⁹ (*resamples*). Pojmem „resample“ se rozumí výběrový soubor (získaný s nahrazením) o stejné velikosti jako soubor normativní. Obrázek

³⁹ Dle Efrona a Tibshiraniho (1993) je 50 bootstrapových resamplů mnohdy dostatečných k poskytnutí přesných odhadů výběrové chyby dané statistiky. S každým nárůstem resamplů se snižuje nárůst zpřesnění odhadů. V normování WJ IV se vycházelo z 250 resamplů, aby byla zajištěna vysoká míra spolehlivosti odhadů výběrových chyb.

3-4 znázorňuje, jak bylo toto „převzorkování“ provedeno v případě testu Rozpoznávání písmen a slov v normativní studii WJ III NU. Představte si, že jsou skóre a věk každého z 8 648 participantů WJ III NU (7 416 v případě původní verze WJ IV) natištěny na pingpongový míček a vloženy do losovacího zařízení, jež je v Obrázku 3-4 znázorněno válcem. Jeden z míčků (tj. participantů v normativní studii) je náhodně vylosován. Jeho věk a skóre jsou zaznamenány a míček navrácen do zařízení (tj. je nahrazen) ještě před dalším losováním. Tento postup je zopakován 8 648krát, z čehož vznikne jeden *resample*. Participant z každého z nich je umístěn do bloků o 50 osobách seřazených dle věku, z čehož je následně obdobně jako u tradiční metody z WJ III vypočítán mediánový věk a W-skóre. Celý proces je zopakován 250krát. Obrázky 3-4 a 3-5 tyto kroky shrnují.

Obrázek 3-5.

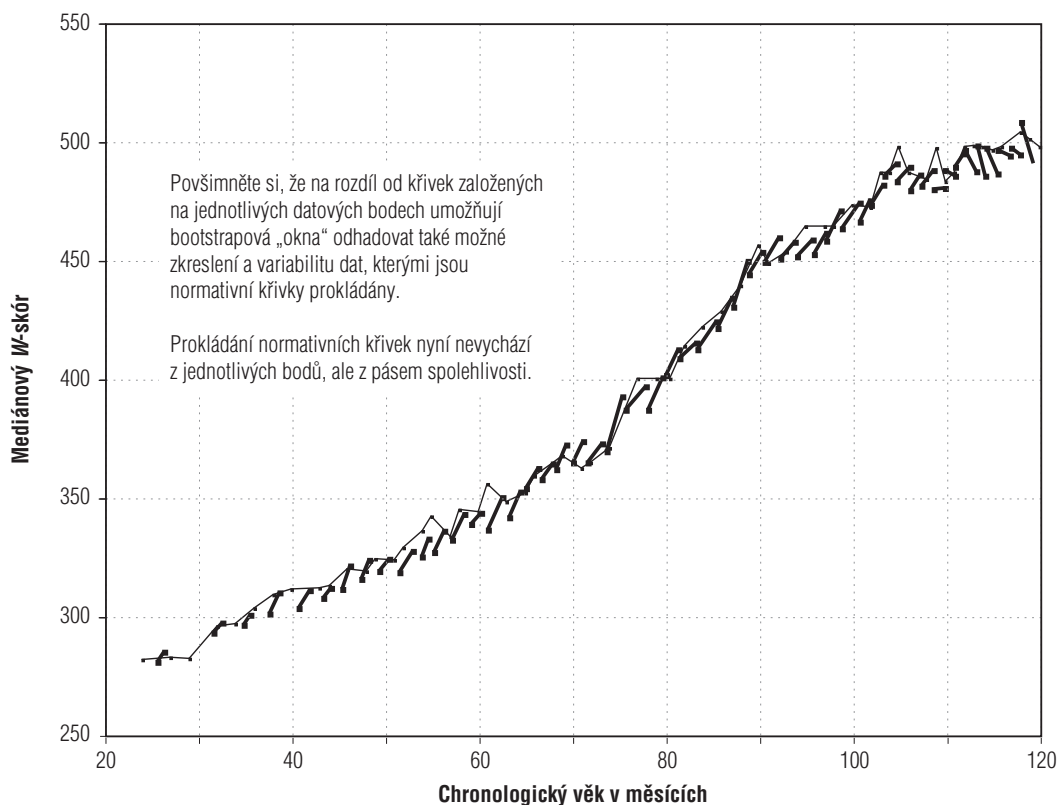
Výpočet pásem spolehlivosti (25. až 75. percentil) pro výběrové statistické indexy vygenerované bootstrapovou metodou (viz Obrázek 3-4).



V této fázi existuje 250 spárovaných hodnot mediánového věku a mediánového W-skóru, jež vzešly z každé skupiny z *resamplu* (viz Obrázek 3-5). Z rozložení mediánových W-skórů v každé z věkových skupin jsou dopočítány 25. a 75. percentil; toto rozpětí reprezentuje středních 50 % vygenerovaných statistik vzorku. Takové okno či pásmo poskytuje empirický odhad míry jistoty statistik výběru, z nichž vychází proces tvorby normativních křivek. Je-li křivka vykreslována z bodu definovaného 25. percentilem chronologického věku a 25. percentilem mediánového W-skóru až k bodu definovanému 75. percentilem těchto statistik, je výsledkem pro každý blok „okno“, skrze něž jsou křivky prokládány. To znamená, že namísto prokládání normativních křivek datovými body jako na Obrázku 3-3 jsou nyní tyto křivky prokládány okny pásem spolehlivosti, jak ukazuje Obrázek 3-6.

Obrázek 3-6.

Srovnání hrubých datových bodů REF W z testu Rozpoznávání písmen a slov z WJ III a bootstrapových „tyčí“ nebo „oken“ z WJ III NU.



Poznámka: Horní a dolní strany „tyčí“ jsou mírně posunuty, aby mohly být zobrazeny ve stejném grafu.

Krom tvorby bootstrapovaných *resamplů* se proces výpočtu norem pro WJ III NU a WJ IV oproti WJ III liší rovněž ve využití vážení jednotlivých účastníků normativní studie. Postupy v těchto nových testech nevyžadují, aby měl každý participant stejnou pravděpodobnost, že bude do studie vybrán. Váhy participantů byly převedeny na pravděpodobnosti výběru, a to tak, aby osoby s vyšší vahou byli pravděpodobněji vybráni a zařazeni do jednotlivých *resamplů*. Někteří účastníci byly vybráni mnohem častěji než jiní, což umožnilo vyvážit demografické charakteristiky vzorku vzhledem k referenční populaci.

Výhody novějších metod výpočtu hodnot REF W

Postup využitý ve WJ III NU a WJ IV má několik zjevných výhod. Zaprvé je zjednodušen výpočet statistik pro každý blok v *resamplu*, protože váhy jednotlivých participantů nyní nejsou součástí kalkulací. Namísto toho jsou nově začleněny do pravděpodobnosti výběru konkrétní osoby do daného *resamplu*. Uvedený postup usnadňuje výpočet komplexnějších statistik (nad úroveň mediánu W-skórů).

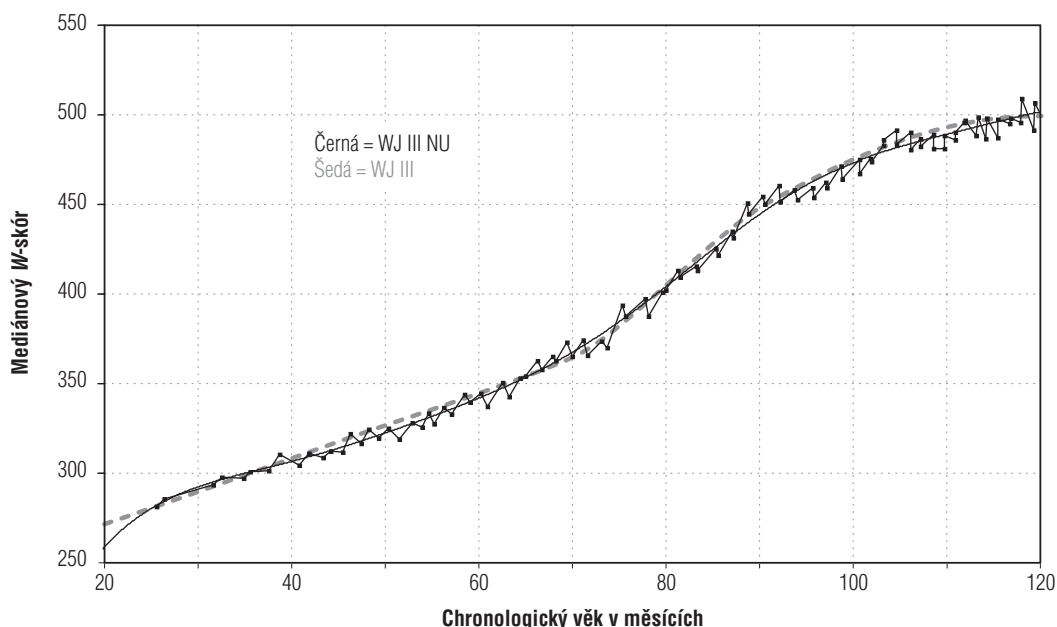
Zadruhé, což je zásadnější, nyní proces prokládání křivek zahrnuje volbu trajektorie skrze řadu pásem spolehlivosti (tyčí či oken) namísto série bodů (viz Obrázek 3-6). V jednotlivých úrovních věku tak dostáváme rozpětí hodnot (se známou mírou jistoty), jež mohou být využity pro získání vyhlazených hodnot REF W. Tradiční postup z WJ, WJ-R či WJ III je snahou volit takové křivky, které se vyhýbají minimu bodů. Volbou průchodu pásmy spolehlivosti je nejistota nevyhnutelně spjatá s procesem tvorby vzorku přiznaná a zjevně pozorovatelná, čímž je omezena tendence normativních křivek „nahánět jednotlivé body“ – praxe ústící v méně přesné normativní křivky. Tato výhoda je patrná na Obrázku 3-6, kde jsou bootstrapová pásma spolehlivosti na příkladu testu z WJ III, Rozpoznávání písmen a slov, kladena přes jednotlivé body z daného testu (z Obrázku 3-2).

Na Obrázku 3-7 jsou ty nejzjevnější rozdíly mezi REF W zobrazené pomocí bodů a pásem patrné mezi věky 35 až 65 měsíců. Pásma spolehlivosti jsou konzistentně nižší než odpovídající body, což napovídá tomu, že jsou hodnoty REF W z WJ III v tomto věkovém rozpětí zkresleny směrem nahoru. Normativní křivka proložená těmito body bude rovněž mít tendenci růst, což vyústí v méně přesné

odhady REF W. Ačkoli se taková křivka na Obrázku 3-3⁴⁰ jevila jako optimální řešení, ve srovnání s řešením vzešlým z pásem spolehlivosti (Obrázek 3-7) je zjevné, že v uvedeném věkovém rozpětí vede výše. Na základě kvantifikace variability v rozsahu odhadů v těchto specifických věkových blocích (viz zobrazené tyče či okna) se potom jako přesnější jeví křivka, která vede níže. Zmíněné křivky jsou rovněž zřetelně rozdílné od zhruba 20 do 30 měsíců věku. V tomto případě odhady založené na bodech ve WJ III vedou ke křivce zjevně posazené výše než ve WJ III NU.

Obrázek 3-7.

Srovnání možných normativních křivek subtestu Rozpoznávání písmen a slov z WJ III (šedě) a WJ III NU (černě).



Poznámka: Propojené datové body jsou ve skutečnosti spojené bootstrapové „tyče“ nebo „okna“ (viz nespojené tyče nebo okna z Obrázku 3-6).

Mezi psychometrickými působícími v praxi je známo, že nejobtížnější je normativní křivky prokládat u nejmladších a nejstarších skupin, protože v těchto věkových kategoriích postrádáme data pod nejmladším a nad nejstarším věkem. Jsou-li křivky prokládány sadou spojitých dat obklopených ostatními body (např. body mezi 60. a 80. měsícem na Obrázku 3-2), využívají příslušné algoritmy také informaci z bezprostředně předešlých či následujících bodů. Oproti tomu, jak vidno na Obrázku 3-2, nemá bod ve 24 měsících žádné předcházející případy, z nichž by bylo možné dovodit, jak jím má křivka procházet. Obdobně (byť nezobrazen na Obrázku 3-2) je výzvou opačný konec spektra, protože za ním již neexistují empirická data, jež by prokládání mohla vypomoci. Výsledkem je značná nejistota, jež obklopuje hodnoty statistik vzorku a tvar normativní křivky u nejmladších a nejstarších osob. V praxi psychometricky typicky extrapolují normativní křivky kousek za extrémní, empiricky získané body, což způsobuje, že je zde diskutovaná činnost na uvedených pólech více uměním než vědou. Jsou-li ovšem, jak je ukázáno na Obrázku 3-7, jako zdrojová data použita bootstrapem získaná pásma spolehlivosti, stávají se obecné trendy na extrémech zjevnějšími. V případě testu WJ III, Rozpoznávání písmen a slov, představeného na Obrázku 3-7, je větší jistota vkládána do řešení, jež je mezi 20 a 36 měsíci posazeno níže.

Obě ilustrativní křivky z Obrázku 3-7 demonstrují, že přístup založený na bootstrapu přináší v tvorbě normativních křivek metodologický pokrok. Takto získaným výsledkům je oproti tradičním postupům z předchozích vydání WJ možno více věřit⁴¹.

Výpočet věkových ekvivalentních skóre⁴²

Pro účely věkově ekvivalentních skóre WJ IV ve vyhodnocovací aplikaci byly použity vyhlazené křivky získané pomocí knihovny cNORM (Lenhard & Lenhard, 2019) ve statistickém software R

40 Adekvátní fit dokládají též příslušné statistiky fitu spojené s řešením z Obrázku 3-3.

41 A podobně je možné více věřit i výsledkům použitým námi – polynomičkému vyhlazování.

42 Přestože používáme analogické postupy, v této části píšeme výhradně o tvorbě českých norem, protože poznámky k odlišnostem postupů by byly příliš rozsáhlé a čtenářsky nezajímavé. V další, poslední části této kapitoly je opět použit překlad amerického originálu, protože postupy použité v české vyhodnocovací aplikaci byly analogické s americkými.

(R Core Team, 2019). Pro každý W-skór (odpovídající jednotlivým hrubým skóřům) byla zjištěna hodnota věku (ve letech a měsících), pro kterou daný W-skór odpovídal hodnotě IQ = 100. Tyto věkové ekvivalenty byly hledány pouze v částech vývojových křivek, ve kterých docházelo k markantnímu růstu. Jakmile křivka dosáhla plateau, hledání věkového ekvivalentu ztratilo smysl.

Výpočet percentilových pořadí a standardních skóřů

I hodnoty percentilových pořadí a standardních skóřů byly založeny na vyhlazených křivkách zmíněných v předchozím odstavci. Postup jejich získávání umožnil vypočítat vyhlazenou křivku pro každou hodnotu věku. Díky tomu jsme pro každou hodnotu věku mohli získat standardizované skóřy a jim odpovídající percentily⁴³.

Výpočet rozdílových skóřů WJ IV

Rozdílové skóřy svému uživateli umožňují činit predikce založené na datech, stejně jako srovnání mezi vybranými testy a klastrovými skóřy odvozenými z baterie WJ IV. Ty lze potom využít při popisu výkonových vzorců, jež mohou být přínosné při diagnostickém rozhodování a plánování výuky. Následují dvě nejčastější použití rozdílových skóřů v diagnostické praxi.

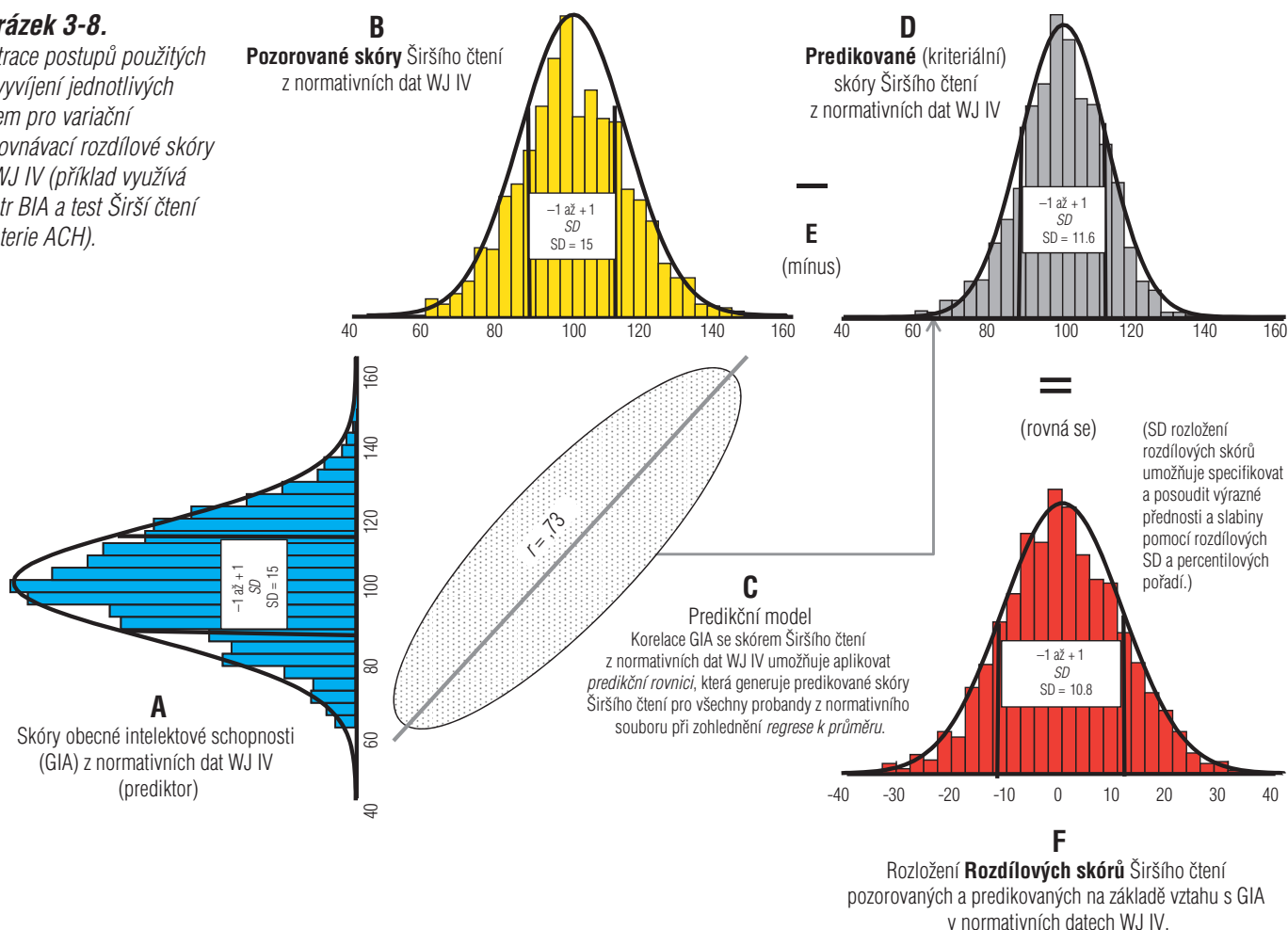
1. K určení toho, zda je relativní pozice osoby v rámci skupiny v testu či klastru (např. Číselných řad ve WJ IV COG) statisticky významně odlišná od probandova relativního postavení v dané skupině v jiném testu či klastru (např. ve Vizualizaci z WJ IV COG).
2. K určení toho, zda je relativní pozice osoby v rámci skupiny v testu či klastru statisticky významně odlišná od očekávání či predikce učiněné na základě jeho či jejího výsledku v „predikujícím“ testu či klastru.

První příklad představuje *rozíl v profilu standardních skóřů/percentilových pořadí*. Druhý příklad vychází z rozložení skutečných rozílů mezi predikujícími a kritériálními skóřy pozorovanými ve vzorku normativní studie. Vývoj norem tohoto typu rozdílových skóřů je popsán dále.

⁴³ Autoři americké verze počítali s odlišnou SD pod mediánem a nad mediánem jednotlivých skóřů. To jsme my nedělali, především proto, že jsme neměli dostatek dat na to, abychom v jednotlivých věkových podskupinách mohli ověřovat šikmost rozložení.

Obrázek 3-8.

Ilustrace postupů použitých při vyvíjení jednotlivých norem pro variační a srovnávací rozdílové skóry ve WJ IV (příklad využívá klast. BIA a test Širšího čtení z baterie ACH).



Poznámka: Všechna rozložení skóř představují skutečné skóry probandů ve věku 9–13 let z konormativního souboru WJ IV. Reálné predikční modely se liší dle věku a ročníku (vývojově se měnící predikční modely).
SD rozložení predikovaných a rozdílových skóř se nerovná 15. To by platilo jen za předpokladu, že by korelace skóř GIA a Širšího čtení byla dokonalá (1).

Výhodou konormování (tvoření společných norem) baterií WJ IV COG, WJ IV OL a WJ IV ACH je možnost výpočtu skutečných rozdílů mezi predikujícími a kriteriálními proměnnými u každé z osob v normativním vzorku, což dále umožňuje zmíněné rozdíly modelovat pro populaci. Ve WJ IV má tento typ rozdílových skóřů dvě podoby: *variance a porovnání*. Obě uvedené formy jsou založeny na společném statistickém modelu, jenž je vizuálně demonstrován na Obrázku 3-8. Rozdíl mezi nimi tkví v tom, který ze skóřů modelu slouží jako prediktor (A na Obrázku 3-8). Variance vychází z predikujícího skóru, jenž je průměrem (nekriteriálních) skóřů z banky testu, která vylučuje kriteriální proměnnou. Porovnání jsou založena na jediném prediktoru, jako je např. klastrový skóř GIA či Gf-Gc. Skórovací algoritmy pro tyto formy rozdílových skóřů byly konstruovány pomocí procedur založených na regresi, podobných výše popsaným postupům u diskrepancí⁴⁴ z WJ-R a WJ III (McGrew et al., 1991; McGrew a Woodcock, 2001).

Postupy vyobrazené na Obrázku 3-8 jsou shodné s vývojem všech srovnávacích a variačních procedur ve WJ IV. Vždy je za užití predikujícího skóru (A) generován cílový, predikovaný (kriteriální), skóř (D), jenž je založen na korelaci mezi predikujícími (A) a skutečně získanými skóry (B) u participantů z normativního vzorku WJ IV. V uvedeném příkladu je korelace mezi predikujícím (A) a získaným (B) skórem rovna 0,73, ve skutečnosti se však korelace vývojově posouvají (dle věku či ročníku) na základě komplexních statistických predikčních modelů založených na polynomicke regresi⁴⁵ (C).

⁴⁴ Termínu „diskrepanční skóř“ zde neuvádíme, protože má i nadbytečný význam jako výsledek původního modelu „diskrepance mezi schopnostmi a výsledky“, jenž se používá při identifikaci specifických poruch učení (SPU). Diskrepanční skóř sice je formou skóru rozdílového, druhý jmenovaný se však liší v metodě výpočtu, účelu a interpretaci.

⁴⁵ Prediktivní modely založené na regresi zohledňující regresi k průměru jsou podobné, jež byly dříve popsány pro diskrepance ve WJ-R a WJ III (McGrew et al., 1991; McGrew a Woodcock, 2001). V případě českého normativního souboru, který byl mnohem méně rozsáhlý než americký, byla použita lineární regrese.

Skutečné skóry (B) mínus skóry predikované (D) vytváří rozdílové skóry (F) pro každého z účastníků normativní studie. V příkladu z Obrázku 3-8, je GIA skór (A), v této věkové skupině se Širším čtením⁴⁶ (B) korelovaný 0,73, vložen do regresního prediktivního modelu (C) s cílem generovat predikované skóry Širšího čtení (D) pro každého z účastníků normativní studie WJ IV v daném věkovém segmentu. Takto predikované skóry (D) jsou pak odečteny od těch skutečně pozorovaných (B), čímž je získáno rozložení rozdílových („skutečnost mínus odhad“) skórů (F) dané skupiny.

Rozložení uvedených skórů v normativním vzorku (F) poskytlo data pro výpočet rozdílových směrodatných odchylek (SD) a percentilových pořadí (PR) za použití postupů konstrukce norem, jež jsou popsány výše. Ve vyhodnocovací aplikaci jsou nazvány jako diskrepanční SD a diskrepanční PR. Diskrepanční SD skór je standardizovaným z-skórem, jenž v jednotkách směrodatných odchylek udává rozdíl mezi rozdílovým skórem osoby a průměrných rozdílovým skórem pro účastníky normativní studie o shodném věku či ročníku, kteří dosáhli stejného predikujícího skóru. Zmíněná SD je použita pro určení statistické významnosti uvedeného rozdílu. Diskrepanční PR udává informaci o poloze probanda v rozložení rozdílových skórů, jichž dosáhly osoby z normativní studie odpovídajícího věku či ročníku.

Je nutno rozeznávat, že zatímco predikující (A) a skutečné standardní (B) skóry zaujímají rozložení o průměru 100 a SD 15, rozložení predikovaných (D) a rozdílových (F) skórů tuto vlastnost nemají. Jak ilustrujeme na příkladu Obrázku 3-9, průměr rozložení predikovaných skórů (D) Širšího čtení je 100, avšak SD je rovna 11,6 (nikoliv 15). Uvedená SD reflektuje důsledek regrese k průměru způsobené nedokonalou korelací mezi rozložením predikovaných (A) a skutečných (B) skórů. SD v rozložení D by byla rovna 15 právě v případě, kdy by predikující (A) a pozorované (B) skóry byly korelovány perfektně ($r = 1$). To znamená, že by uživatelé neměli predikované skóry nabídnuté vyhodnocovací aplikací interpretovat tak, že mají shodný význam jako skóry standardní, mající průměr 100 a odchylku 15. Pokud by v příkladu z Obrázku 3-8 examinátor interpretoval výsledek 85 jako ležící jednu směrodatnou odchylku pod průměrem, výrazně by tím nadhodnotil skutečně predikovaných skór, z něž plyne, že výsledek 85 odpovídá 1,29 SD pod průměrem ($100 - 85 = 15$; $15/11,6 = 1,29$ SD).

Rozložení rozdílových skórů (F) pak má vždy průměr 0 a unikátní SD, která obecně není rovna 15. Na Obrázku 3-8 je za předpokladu korelace 0,73 mezi skóry v GIA a Širším čtení a věkové skupiny mezi 9 a 13 lety SD daného rozložení rovna 10,8. Rozhodne-li se examinátor, že má rozdílový skór být menší nebo roven -1,5 SD, aby byl brán za významný pro diagnostické a klasifikační účely, pak by bylo zapotřebí získat alespoň rozdílový skór přibližně -17,5 ($10,8 \times -1,5 = -17,5$). Domnívá-li se examinátor mylně, že SD činí 15 bodů (a požadovaný rozdíl -1,5 SD je tedy roven zhruba -22 či -23 bodům), pak se mu či jí nepodaří identifikovat existující signifikantní rozdíl. Vyhodnocovací aplikace potenciál pro takovou chybu odstraňuje tím, že sama určuje, zda je srovnávací či variační rozdílový skór na stanovené úrovni významný (např. 1,5 SD).

⁴⁶ Jde o klastrový skór, který je součástí americké baterie WJ Testy školních dovedností, která nebyla převedena do češtiny. Jde o klastr obsahující šířeji chápané čtenářské dovednosti, jako je např. porozumění psanému textu a plynulost čtení. Zde je použit proto, že jsme z americké verze Technického manuálu převzali výborně graficky zpracované vysvětlení, jak se pracuje s prediktivními modely, které zde vysvětlujeme.

Reliabilita⁴⁷

Reliabilita vypovídá o přesnosti testového skóru. Vysoká reliabilita indikuje, že naměřená hodnota probanda by se pravděpodobně příliš nezměnila, pokud bychom ho otestovali znovu za podobných podmínek. Reliabilita je nutnou, ale nikoli dostačující podmínkou validity. Vysoká reliabilita nutně neznamená, že testový skór je s ohledem na svůj účel validní, nicméně důkazy o reliabilitě představují důležitý prvek při dokazování validity testu. Tato kapitola (a s ní související Přílohy B a C) obsahuje důkazy demonstrující, že testové a klastrové skóry testu Woodcock-Johnson IV (WJ IV) (Schrang, McGrew & Mather, 2014a) jsou dostatečně reliabilní a přesné pro měření kognitivních a jazykových schopností a školních dovedností pro většinu věkového spektra. Kapitola nejprve shrnuje některé základní koncepty reliability a chyby měření. Následně popisuje několik různých typů analýz reliability provedených pro testové a klastrové skóry WJ IV a uvádí jejich výsledky⁴⁸.

Koncepty reliability

Protože žádný test nelze považovat za dokonalou míru nějakého rysu, všechny testové skóry jsou do určité míry zatíženy chybou. Předpokládáme, že tato chybová komponenta je nesystematická a náhodně rozložená okolo nuly. V klasické testové teorii (CTT) (de Ayala, 2009; Raykov & Marcoulides, 2011; česky Urbánek, Denglerová, Širůček, 2011) je konkrétní pozorovaný skór chápán jako složený ze dvou částí: pravého skóru, tedy skóru, který bychom obdrželi po změření hypotetickým „perfektním“ testem, a chybové komponenty:

$$X_{\text{pozorovaný}} = X_{\text{pravý}} + X_{\text{chybový}} , \quad (4.1)$$

kde $X_{\text{pravý}}$ a $X_{\text{chybový}}$ jsou nezávislé a nekorelované. Pokud máme k dispozici velké množství individuálních skórů, mohli bychom vztah mezi průměry jejich distribucí popsat následujícím způsobem:

$$\bar{X}_{\text{pozorovaný}} = \bar{X}_{\text{pravý}} + \bar{X}_{\text{chybový}} , \quad (4.2)$$

kde $\bar{X}_{\text{chybový}}$ je roven nule. Vztah mezi rozptyly vypadá takto:

$$SD_{\text{pozorovaný}}^2 = SD_{\text{pravý}}^2 + SD_{\text{chybový}}^2 . \quad (4.3)$$

Reliabilitu (r_{11}) testu definujeme jako podíl rozptylu pravých skórů a rozptylu pozorovaných skórů:

$$r_{11} = \frac{SD_{\text{pravý}}^2}{SD_{\text{pozorovaný}}^2} , \quad (4.4)$$

nebo ekvivalentně:

$$r_{11} = 1 - \frac{SD_{\text{chybový}}^2}{SD_{\text{pozorovaný}}^2} . \quad (4.5)$$

Poslední uvedená rovnice popisuje testovou reliabilitu vyjádřenou rozptyly pozorovaných a chybových skórů.

⁴⁷ Autorkou překladu této části je Edita Chvojková.

⁴⁸ Výsledky v této kapitole jsou limitovány na reliabilitu WJ IV COG a vzhledem k novosti testu v českém prostředí zjištěnou pouze jedním způsobem.

Odmocninu z chybového rozptylu v rovnici 4.5 (tedy $SD_{chybový}^2$ nebo také směrodatná odchylka (SD) rozdílů mezi pozorovanými a pravými skóry) nazýváme standardní chyba měření (angl. *standard error of measurement*, *SEM*). Dosazením SEM do rovnice 4.5 dostáváme

$$r_{11} = 1 - \frac{SEM^2}{SD_{pozorovaný}^2} . \quad (4.6)$$

SEM se často využívá jako ukazatel přesnosti konkrétního individuálního skóru nebo také přesnosti, s níž může být odpovídající pravý skóre lokalizován na škále. Mezi reliabilitou a SEM existuje inverzní vztah; s klesající SEM roste reliabilita. Chyby měření testů WJ IV mohou být vypočteny v jednotkách W-skórů podle tradičního vzorce popisujícího vztah koeficientu reliability a směrodatné odchylky souboru testových skórů:

$$SEM = SD_{pozorovaný} \sqrt{1 - r_{11}} . \quad (4.7)$$

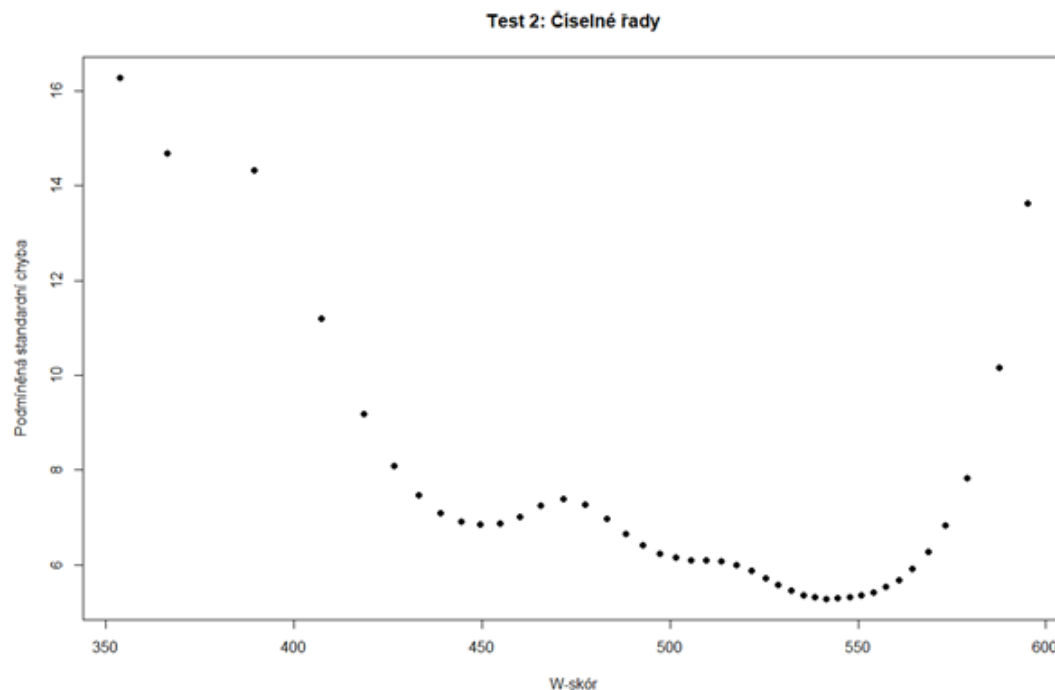
Tyto standardní chyby měření (SEM) jsou skupinové statistiky a lze je interpretovat jako průměry chyb měření individuálních W-skórů v daném věku či ročníku.

Na rozdíl od odhadů reliability v rámci klasické testové teorie (CTT), které při měření schopností předpokládají stejnou přesnost napříč celým jejím rozsahem, na teorii odpovědi na položku (IRT; de Ayala 2009; Raykov & Marcoulides, 2011) založené raschovské procedury vedoucí k odvození W-skórů umožňují jedinečný odhad chyby měření každého skóru schopnosti spojeného s konkrétním hrubým skórem. Použití Raschova modelu a škály W tedy umožňují uživatelům testu přímo srovnávat hodnoty SEM všech testů na společné škále na všech úrovních. Pokud je například SEM jednoho testu na věkové úrovni 5 let 6,0 a druhého testu ve věku 8 let 3,0, můžeme říci, že skóry druhého testu ve věku 8 let jsou dvakrát přesnější než skóry testu prvního ve věku 5 let. Při uniformní hustotě položek v testu (tj. stejných rozestupech položek na škále W dle obtížnosti), budou odhady SEM vyjádřené v jednotkách W nižší v oblasti středu a větší v extrémech. Přesná hodnota SEM v konkrétním pásmu testu je funkcí počtu položek a jejich hustoty v tomto pásmu.

Obrázek 4-1 znázorňuje typický vztah mezi chybami měření a úrovní schopnosti na škále W v celém rozsahu testu. Tečky v grafu reprezentují hrubé skóry ve Woodcock-Johnson IV Testech kognitivních schopností (WJ IV COG) (Schrang, McGrew, & Mather, 2014b), ze Standardní baterie Test 2: Číselné řady; se začátkem v 0 (W-skóre 353,66) a maximálním skórem 42 (W-skóre 595,05). Povšimněme si, že chyba měření rapidně narůstá na koncích rozložení hrubých skórů. Rovněž věnujme pozornost faktu, že hrubé skóry jsou mezi W -hodnotami 497,29 (hrubý skóre 18) a 525,36 (hrubý skóre 25) od sebe vzdáleny téměř identicky. Toto stejnoměrné rozložení výsledných W -schopností je důsledkem výběru položek s relativně uniformním odstupňováním W -obtížností v rozsahu celého testu. Pokud mají položky uniformní odstupňování W -schopností, mají hrubé skóry a W -skóry přibližně lineární vztah, s výjimkou spodních a horních extrémů. Tento vztah je obzvláště patrný u testů s širokým rozpětím hodnot, jako jsou ty, které jsou součástí baterie WJ IV.

Obrázek 4-1.

Vztah chyby měření
a W-skóru pro Test 2:
Číselné řady



Koeficienty reliability

O koeficientu reliability můžeme přemýšlet jako o indexu přesnosti, s níž odhadujeme, jak si daný člověk stojí v rámci určité skupiny, či kolikátý je v pořadí. Z vysoké hodnoty koeficientu reliability vyplývá, že skóre člověka nebo jeho pořadí ve skupině by byly podobné při opakovaných administracích testu. Pokud podrobně rozebereme jednotlivé části rovnice 4.6, vidíme, že vysoká reliability vyplývá jednak z malé standardní chyby měření (tj. dobré přesnosti skóre), nebo z vysoké hodnoty rozptylu pozorovaných skóre ($SD_{\text{pozorovány}}^2$; tedy větší variability testových skóre v rámci skupiny). Naopak větší hodnoty SEM a/nebo omezená variabilita testových skóre v rámci skupiny může vést k podhodnocení velikosti koeficientu reliability.

Může se přihodit, že test produkuje skóre s dobrou přesností, což se odráží v malé SEM, ale pokud je rozsah hodnot měřeného rysu omezený, přesnost relativního postavení ve skupině je nízká. Omezené rozpětí skóre a s ním spojenou nízkou r_{11} můžeme pozorovat, pokud pro účely analýz reliability zvolíme skupinu, jejíž data mají malé rozpětí skóre (např. skupinu s převážně nulovými nebo velmi nízkými skóre nebo skupinu s většinou perfektními a velmi vysokými skóre), nebo tehdy, kdy měřený rys disponuje jen omezenou variabilitou v subpopulaci na konkrétní úrovni schopnosti (McGrew, Schrank, & Woodcock, 2007).

V případech, kdy je nízká reliability způsobena omezenou variabilitou ve výběru nebo populaci, je tvrzení, že jde o „nereliabilní“ (a tedy nekvalitní) nástroj analogické výroku, že konkrétní skládací metr není spolehlivým nástrojem měření, protože všechny předměty, které měříme, jsou přibližně 91 centimetrů dlouhé (McGrew et al., 2007). Tabulka 4-1 ilustruje tuto problematiku na konkrétním příkladu z WJ IV. Tato tabulka obsahuje informace o SEM, směrodatných odchylkách a reliabilitách pro věkové skupiny 11 let a 50-59 let pro Woodcock Johnson IV Testy kognitivních schopností (WJ IV COG) (Schrank, McGrew, & Mather, 2014b), Test 5: Fonologické zpracování⁴⁹. Pověšme si, že koeficient reliability je pro věkovou skupinu 11letých relativně nízký (0,78) v porovnání se skupinou 50-59 let (0,90). Podrobnější prostudování tabulky 4-1 odhalí, že přesnost testových skóre vyjádřená SEM je pro obě skupiny přibližně stejná (4,60, resp. 4,58). Rozdíly v reliabilitě jsou tedy důsledkem většího rozptylu skóre schopnosti (SD) měřené tímto testem ve skupině 50-59 let (14,49), v porovnání s rozptylem skóre schopnosti v 11 letech (9,81). Test je tedy stejně přesný ve skupině 11letých

49 V tomto případě jde o příklad založený na amerických datech.

děti i ve skupině padesátníků, ale zdánlivá přesnost relativních pozic v 11 letech není tak dobrá, protože v této věkové skupině je menší rozsah individuálních rozdílů. Relativně nízký koeficient reliability je tedy spíše funkcí výběru, nikoli testu samotného.

Tabulka 4-1.

Vybrané ukazatele reliability
z Testu 5 WJ IV COG:
Fonologické zpracování

Statistika	Věk 11	Věk 50–59
SEM	4,60	4,58
SD	9,81	14,49
r_{11}	0,78	0,90

Pozn.: Tato tabulka je založena na amerických datech.

Reliability testů

Statistiky reliability byly vypočteny pro všechny testy z baterie WJ IV, a to ve všech věkových skupinách, v nichž má být test používán. Výpočty reliability zahrnovaly všechna data ze standardizačního vzorku ze všech věkových kohort⁵⁰.

Standardní chyba koeficientů reliability umožňuje výpočet intervalu spolehlivosti, v němž se s předem zvolenou pravděpodobností nachází skutečný koeficient reliability. V Tabulce 4-2 jsou uvedeny 68% intervaly spolehlivosti pro několik typických hodnot koeficientů reliability a velikosti výběrů. Tyto intervaly byly určeny nejprve transformací reliabilit na hodnoty Fisherova z' -rozložení, jak jsou uvedeny ve statistických tabulkách, standardní chyby z -hodnot pak byly vypočteny z následující rovnice:

$$SE_{z'} = \sqrt{\frac{1}{N-3}} \quad (4.8)$$

Tyto směrodatné chyby byly využity ke konstrukci intervalů spolehlivosti z' -hodnot. Nakonec byly transformovány zpět na škálu reliabilit, odpovídající hodnotám uvedených v tabulce 4-2.

Tabulka 4-2.

Intervaly spolehlivosti
v rozpětí ± 1 SEM pro
vybrané hodnoty reliability
a velikosti výběru

Reliabilita	Velikosti výběru		
	100	200	300
0,80	0,760–0,835	0,773–0,824	0,782–0,815
0,90	0,879–0,918	0,886–0,912	0,890–0,909
0,95	0,939–0,959	0,942–0,957	0,945–0,954

Testy a subtesty s dichotomicky skórovanými položkami

Reliability ve smyslu vnitřní konzistence pro všechny testy bez rychlostní složky a subtesty s dichotomicky skórovanými položkami byly vypočteny přístupem split-half⁵¹. Vypočítali jsme hrubé skóre respondentů ze standardizačního výběru pro sudé a liché položky těchto testů. Tyto dvě sady položek jsme následně korelovali. Základní pravidla ukončení používání pro účely standardizační studie byla dostatečně přísná, aby pravděpodobnost, že respondent nevyřeší položku jednodušší, než je jeho bazální úroveň, a naopak vyřeší položku nad stropem, byla velmi nízká. Všechny odpovědi pod respondentovou základní úrovní byly tedy považovány za správné a všechny odpovědi nad respondentovým stropem byly skórovány jako nesprávné. Vzniklé koeficienty byly následně korigovány vzhledem k finální délce testu podle Spearman-Brownova vzorce.

Testy s vícebodovým skórováním

Protože přístup split-half není v případě testů a subtestů s rychlostní složkou (např. Fonologické

⁵⁰ Věkové kohorty pro analýzy reliability jsou následující: 5;0 – 6;11, 7;0 – 8;11, 9;0 – 10;11, 11;0 – 12;11, 13;0 – 14;11, 15;0 – 16;11, 17;0 – 18;11, 19;0 – 20;11, 21;0 – 25;11, 26;0 – 30;11, 31;0 – 40;11, 41;0 – 50;11, 51;0 – 60;11, 61;0 – 70;11, 71;0 – 80;11.

⁵¹ Koeficienty reliability v české standardizační studii byly vypočteny na základě výše uvedeného vzorce (4.6).

zpracování – Slovní fluence) vhodný⁵², a stejně tak je tomu u testů obsahujících vícebodové skórování (např. některé z testů WJ IV ACH), byly reliability pro tyto testy odvozeny z informací poskytnutých Raschovým modelem. Raschův model stojící za škálou W poskytuje standardní chybu měření (SEM) odhadu schopnosti každého respondenta ve standardizačním vzorku. Rozptyly pozorovaných skóre ($SD_{\text{pozorovaný}}^2$) a hodnoty středních kvadratických chyb (angl. *mean-square error*) byly vypočteny pro všechny respondenty v daných věkových kohortách a dosazeny do rovnice 4.6 pro výpočet reliability.

Použití středních kvadratických chyb odůvodňujeme následujícím postupem. Rovnici 4.6 uvádíme pro usnadnění ještě jednou:

$$r_{11} = 1 - \frac{SEM^2}{SD_{\text{pozorovaný}}^2}, \quad (4.9)$$

kde SEM je ekvivalentní členu SD_{chyba} , jak jsme diskutovali výše. Nyní vezměme v úvahu obecný vzorec pro výpočet rozptylu:

$$SD_Y^2 = \frac{N\Sigma Y^2 - (\Sigma Y)^2}{N(N-1)}. \quad (4.10)$$

Pokud určíme, že Y bude chybový člen, pak s ohledem na to, že chyby jsou náhodně rozloženy kolem nuly, je jejich suma rovněž rovna nule. Pak

$$SD_Y^2 = \frac{\Sigma Y^2}{(N-1)}, \quad (4.11)$$

což je hodnota, která se blíží hodnotě střední kvadratické chyby (angl. *mean-squared error*), pokud velikost výběru N nabývá dostatečné velikosti. Rozdíly mezi rozptyly chybových skóre a střední kvadratickou chybou (angl. *mean-squared error*) jsou dostatečně malé, abychom je při výpočtu těchto reliability mohli zanedbat. Střední kvadratická chyba jako čítec zlomku ve vzorci pro reliability je po vynásobení výrazem $(N-1)/N$ ekvivalentní rozptylu chybových skóre.

Testy se subtesty

Některé testy WJ IV COG jako Fonologické zpracování, Vizualizace nebo Všeobecné znalosti sestávají ze dvou či tří subtestů. Reliability kompozitních skóre může být vypočtena s použitím následující rovnice (Mosier, 1943):

$$r_{cc} = 1 - \frac{\Sigma SD_j^2 - \Sigma SD_j^2 r_{jj}}{\Sigma SD_j^2 + 2\Sigma SD_j SD_k r_{jk}}, \quad (4.12)$$

kde r_{cc} je reliability složeného skóre, SD_j je směrodatná odchylka skóre schopnosti respondentů v testu j , SD_k je směrodatná odchylka skóre schopnosti respondentů v testu k , r_{jj} je reliability testu j , a r_{jk} je korelace mezi testy j a k . Rovnice 4.12 byla použita k výpočtu nevážených reliability pro testy se subtesty⁵³. Reliability subtestů jsme obdrželi buď metodou split-half nebo výše popsáním raschovským postupem.

Příloha B uvádí koeficienty reliability (r_{cc}) a standardní chyby měření (SEM) pro všechny testy WJ IV ve všech 15 věkových kohortách. Tabulky v příloze B zahrnují také průměry W škál a směrodatné odchylky všech testů.

Z přehledu mediánů reliability pro každý test v Příloze B je patrná míra, do níž jednotlivé testové koeficienty dosahují žádané hodnoty 0,80, či ji přesahují. Všechny 18 reportovaných mediánů reliability je výrazně vyšších než 0,90. Přestože na úrovni individuálních testů se jedná o velmi vysoké koeficienty, pro interpretaci doporučujeme spíše klastrové skóre WJ-IV, a to především v případech, kdy

52 Metody výpočtu reliability založené na pojetí vnitřní konzistence, jako například přístup split-half, předpokládají, že průměrná korelace položek testu je stejná jako průměrná korelace hypotetických paralelních forem vzniklých rozdělením testu na dva kratší testy (např. na sudé a liché položky). Tento předpoklad je porušen, pokud testy obsahují položky, které produkují navzájem odlišná rozpětí skóre (v případě WJ IV testy s vícebodovým skórováním). V takovém případě může split-half procedura vést k vytvoření testů, které nelze považovat za ekvivalentní, protože položky v jedné polovině mohou mít vyšší maximum celkového skóre než položky v druhé polovině.

53 Tato informace se týká amerického standardizačního souboru. V případě českých dat byl i v případě testů skládajících se z více subtestů použit při výpočtu reliability vzorec (4.6), který je se vzorem (4.12) ekvivalentní.

činíme důležitá rozhodnutí na individuální úrovni. Klastrové skóry jsou založeny na kombinaci dvou či více testů, a tudíž dosahují konzistentně vyšších reliabilit.

Testy s rychlostní složkou

Jak jsme popisovali v Kapítolce 2, testy WJ IV s rychlostní složkou (Hledání písmen, Fonologické zpracování – Slovní fluence, Hledání čísel a Vyhledávání dvojic) byly kalibrovány za pomoci metriky založené na rychlosti. Přestože je tento přístup užitečný pro kalibraci položek a seřazení respondentů, přináší nadhodnocené standardní chyby odhadů schopností v důsledku omezeného počtu možných skóreů pro každý časový interval. Z tohoto důvodu také nejsou pro výpočet reliabilit těchto testů vhodné raschovské procedury, které jsme aplikovali na testy s vícebodovou odpovědí. Namísto toho byla u všech rychlostních testů provedena testová-retestová studie. Respondentům ve třech různých věkových skupinách byla administrována standardizační verze testů s rychlostní složkou, následovaná administrací té stejné formy po uplynutí jednoho dne. Interval opakovaného testování byl záměrně co nejvíce zkrácen, aby minimalizoval možnosti změn v testových skórech podmíněné změnami momentálního stavu respondenta či úrovně jeho latentního rysu. Mezi první a druhou administrací jsme spočítali korelaci, a to s korekcí proti omezené variabilitě vzorku (Sackett & Yang, 2000).

Tabulka 4-3 také zahrnuje průměrné rozdíly ve *W*-skórech mezi první a druhou administrací a s nimi spojenou změnu na škále směrodatných odchylek (na základě prvního pokusu). Přestože rozdíly na škále směrodatných odchylek v rámci opakovaného testování jsou často přičítány „efektu zácviku“⁵⁴, tyto relativně velké rozdíly *SD* jsou pravděpodobně způsobeny jednodenním odstupem mezi testem a retestem. Při první administraci bylo všech osm⁵⁵ prověřovaných testů pro respondenty nových, a to zvláště v případě kognitivních testů. V době druhé administrace (pouze o jeden den později) paměť a zkušenost již potlačila prvek novosti, což se promítlo do zvýšení skóreů. Průměrné zlepšení skóreů napříč všemi testy ve třech věkových skupinách bylo 0,36, 0,82 a 0,46 v jednotkách směrodatné odchylky. Uživatele upozorňujeme, aby na základě informací uvedených v Tabulce 4-3 nečinili rozhodnutí o efektu zácviku v případě delšího intervalu mezi testem a retestem⁵⁶ – což je informace, která se často používá k určení toho, jak dlouhý čas je nutný k tomu, aby skóre pravděpodobně nebyl ovlivněn efektem zácviku.

54 APA psychologický slovník (VandenBos, 2007) definuje *efekt zácviku* jako „jakoukoli změnu nebo zlepšení, která je výsledkem tréninku či opakování položek či aktivit“ (s. 719).

55 My uvádíme pouze tři testy z WJ IV Cog.

56 Především proto, že uvedené hodnoty jsou založeny na amerických datech. Přestože tři testy, jejichž výsledky uvádíme, jsou v české verzi zcela totožné s verzí americkou, data o testové-retestové reliabilitě těchto testů zatím nemáme k dispozici. Vliv efektu zácviku při opakovaném testování stejným inteligenčním testem v literatuře široce diskutován (Kaufmann, 2009). Výzkumně podložená doporučení pro praxi udávají, že stejný test by neměl být administrován v době menší než 12 měsíců od první administrace (AAIDD, 2010; Kaufmann & Lichtenberger, 2002). Pokud je nutné administraci opakovat, jsou v případě WJ IV ACH k dispozici paralelní formy.

Tabulka 4-3.

Popisné statistiky
a koeficienty testové-
retestové reliability
z americké studie
rychlostních testů WJ IV

Věk 7–11								
Test	n	Průměr		Směrodatná odchylka		r_{12}	Rozdíl průměrů	Rozdíl v jednotkách SD
		Test	Retest	Test	Retest			
Hledání písmen	47	506,73	513,92	23,55	27,48	0,91	7,19	0,31
Hledání čísel	47	442,59	450,35	36,06	32,93	0,85	7,76	0,22
Vyhledávání dvojic	47	498,03	515,28	29,44	30,61	0,89	17,25	0,59
Věk 14–17								
Test	n	Průměr		Směrodatná odchylka		r_{12}	Rozdíl průměrů	Rozdíl v jednotkách SD
		Test	Retest	Test	Retest			
Hledání písmen	49	554,40	564,41	10,13	11,65	0,88	10,01	0,99
Hledání čísel	49	491,91	496,32	11,20	13,22	0,84	4,41	0,39
Vyhledávání dvojic	49	550,94	563,87	18,33	16,32	0,89	12,93	0,71
Věk 26–79								
Test	n	Průměr		Směrodatná odchylka		r_{12}	Rozdíl průměrů	Rozdíl v jednotkách SD
		Test	Retest	Test	Retest			
Hledání písmen	50	546,95	555,27	18,40	19,30	0,91	8,32	0,45
Hledání čísel	50	489,21	494,98	16,05	16,78	0,88	5,77	0,36
Vyhledávání dvojic	50	542,68	558,19	19,79	20,28	0,95	15,51	0,78

Pozn.: Průměry a směrodatné odchylky jsou uvedeny v metrice W skóru. Rozdíl v jednotkách SD byl vypočten vydělením průměrné změny mezi první a druhou administrací směrodatnou odchylkou první administrace v jednotkách W.

Reliability klastrů

Reliability klastrových skóre WJ IV byly vypočteny Mosierovým (1943) neváženým kompozitem z rovnice 4.12. Vztah mezi reliabilitou klastru, směrodatnou odchylkou a standardní chybou měření může být vyjádřen následovně:

$$r_{cc} = 1 - \frac{SEM_{klast}^2}{SD_{klast}^2} . \quad (4.13)$$

Povšimněte si podobnosti rovnice 4.13 s rovnicí 4.6.

Příloha C uvádí průměry, směrodatné odchylky, reliability klastrových skóre a standardní chyby měření pro klastry WJ IV v rozsahu jejich zamýšleného užití v každé věkové kohortě. Hodnoty SEM jsou uvedeny jak na škále W-skóre, tak standardních skóre (SS). Z přehledu mediánů reliability pro každý klaster je patrné, že všechny jsou výrazně vyšší než 0,90.

Důkazy o validitě WJ IV⁵⁷

První kapitola amerického technického manuálu obsahuje návrhy, jak používat a interpretovat testové skóry baterie Woodcock-Johnson IV (WJ IV) (Schrack, McGrew & Mather, 2014a). Tato kapitola obsahuje několik různých typů důkazů podporujících tato dříve uvedená tvrzení, které jsou prezentovány v souladu se *Standards pro pedagogické a psychologické testování* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement In Education [NCME], 1999, 2001). Zahrnuje rovněž diskuzi o validitě s ohledem na tradiční vymezení potřeby věcné, interní a externí validity (Benson, 1998; Cronbach, 1971; Cronbach & Meehl, 1955; Loevinger, 1957; Messick, 1989; Nunnally, 1978) a přináší také rámec pro popis unikátního přínosu Raschova modelu k argumentům týkajícím se validity (Wolfe a Smith, 2007). Většina důkazů prezentovaných v této kapitole se přímo dotýká poznatků komentovaných v kapitole 1, tento manuál a doprovodná WJ IV Příručka (Mather & Wendling, 2014a, 2014b, 2014c) poskytují podpůrné doklady o validitě interpretace testových skóre WJ IV, protože tyto manuály obsahují důležité informace o vývoji, administraci, skórování a interpretaci WJ IV⁵⁸.

Reprezentativnost obsahu baterie WJ IV a její procesuální a konstruktové pokrytí

WJ IV obsahuje testy měřící komplexní sadu unikátních schopností tvořících intelektové schopnosti, schopnosti související s mluveným jazykem a se školními dovednostmi, to vše v rámci teorie CHC⁵⁹. Důkazy, které tato tvrzení podporují, jsou často zahrnuty pod pojmy „obsahová validita“ či „věcná validita“. Pro skóry WJ IV jsou tyto důkazy poskytovány díky specifikaci obsahu testů, stejně jako jednotlivých klastrů v souladu se současným stavem poznání v oblasti CHC teorie. Tyto argumenty ve prospěch validity WJ IV jsou založeny na teoriích obsažených ve třech starších vydáních baterie: *Woodcock-Johnson Psycho-Educational Battery* (WJ) (Woodcock & Johnson, 1977), *Woodcock-Johnson Psycho-Educational Battery-Revised* (WJ-R) (Woodcock & Johnson, 1989) a *Woodcock-Johnson III* (WJ III) (Woodcock, McGrew, & Mather, 2001; 2007). WJ-R a WJ III jsou založeny na několika na sebe navazujících revizích Cattell-Hornovy rozšířené teorie *Gf-Gc* a Cattell-Horn-Carrollovy (CHC) teorie kognitivních schopností (McGrew, 2005, 2009; Schneider & McGrew, 2012). Braden a Niebling (2012) v nezávislé recenzi posoudili kvalitu důkazů o obsahové validitě WJ III, na níž série WJ nepřestává stavět, jako blízkou pozitivnímu konci jejich hodnotící škály. Ve své recenzi ocenili WJ III celkovým hodnocením 4 v oblasti „kvality důkazů o validitě“, a to na 5-bodové škále, kde 1 bod znamenal důkazy slabé a 5 důkazy silné. Hodnocení bylo založeno na (a) zevrubném popisu CHC teorie na níž je obsah testu založen; (b) použití více různých testů k měření každé z širokých domén schopností CHC; (c) využití teorie odpovědi na položku (IRT), konkrétně Raschova modelu, k zajištění toho, že každý test představuje jednodimenzionální škálu měřící jedinou úzkou schopnost; (d) zahrnutí položek v rozpětí pokrývajícím zpracování na nízké úrovni až k vysokofunkčnímu myšlení a usuzování; (e) přizvání panelu expertů ke zhodnocení obsahu testu; a v neposlední řadě (f) práci devíti recenzentů, jejichž úkolem bylo identifikovat položky, které byly potenciálně zkreslené v neprospěch žen, postižených osob a kulturních či jazykových menšin. Cílem tohoto manuálu je popsat stejné formy důkazů o obsahové validitě pro WJ IV.

57 Autorem překladu většiny této kapitoly je Petr Pališek.

58 I když se většina informací v této kapitole týká primárně americké verze WJ IV, jsou užitečné i pro českou verzi – principy testů obsažených v české verzi jsou stejné nebo analogické, jako v americké verzi, takže se dají realisticky očekávat podobné výsledky i v případě studií s českou verzí baterie.

59 Pro více informací o současné teorii CHC a testech a klastrech WJ IV viz kapitulu 1 a Přílohu A.

Pokrytí teorie CHC

Původní vývojová dokumentace WJ IV se snaží posunout plán testů až „za CHC teorii“ (McGrew, 2012; Schneider & McGrew, 2012), jak byla definována v rámci WJ III. Kromě zohlednění současné teorie CHC, která zastřešuje design celé baterie, byl plán vývoje WJ IV ovlivněn i současnými poznatky z neurokognitivního, neuropsychologického a vývojově-psychologického výzkumu. Popis teoretických pilířů baterie WJ IV lze nalézt v kapitole 1 a Příloze A. Obsah Woodcock-Johnson IV Testů mluveného jazyka (*Woodcock-Johnson Tests of Oral Language*, WJ IV OL) (Schrang, Mather, & McGrew, 2014b) a Woodcock-Johnson IV Testů školních dovedností (*Woodcock-Johnson Tests of Achievement*, WJ IV ACH) (Schrang, Mather, & McGrew, 2014a), včetně jejich klastrových skóřů, byl navržen tak, aby pokryl důležité oblasti kurikula, jazykových kompetencí a školních dovedností, jak je definuje americká federální legislativa.

Rozdíl mezi širokými a úzkými schopnostmi představuje v rámci CHC teorie velmi důležitý koncept. V designu WJ III hrála klíčovou roli a prosadila se rovněž v designu a interpretacích většiny dalších důležitých baterií pro měření intelektu (Flanagan, Alfonso, & Ortiz, 2012; Flanagan, Ortiz, & Alfonso, 2013; Keith & Reynolds, 2010; Schneider & McGrew, 2012). Stejně jako v případě WJ III byla většina testů WJ IV navržena tak, aby měřila vždy jedinou úzkou schopnost. Tento přístup, jak plánovat vývoj testu, založený na CHC, který byl poprvé použit při operacionalizaci WJ III, se soustředí na zvýšení reprezentativnosti obsahu konstruktů CHC a na eliminaci rozptylu, který je vzhledem k těmto konstruktům irelevantní (Benson, 1998; McGrew & Flanagan, 1998; Messick, 1995). Za účelem rozšíření byly klastery WJ III konstruovány tak, aby pokryly aspoň dvě kvalitativně odlišné úzké schopnosti. Princip interpretace založený na klastrech byl zaveden proto, aby se zvýšila obsahová validita nástrojů měřících široké schopnosti, jakými jsou čtení, fluidní usuzování nebo obecná inteligence.

Vývojová dokumentace WJ IV je v mnoha ohledech podobná předchozí verzi pro WJ III, významně se však odlišuje v základních akcentech. Princip WJ III vedoucí ke konstrukci relativně čistých, širokých, na CHC založených kognitivních klastrů byl v designu WJ IV zachován, organizace testových příruček, zejména dvou pro testy kognitivních schopností (*Woodcock-Johnson Tests of Cognitive Abilities*, WJ IV COG; Schrang, McGrew, & Mather, 2014b), je nicméně postavena na tom, čemu McGrew říká „hybridně široký plus úzký“ přístup k organizaci intelektové baterie. Ten je založen na přehledu více než 20 let trvajícím výzkumu v oblasti CHC týkajícího se vztahu kognitivních schopností s akademickým výkonem (McGrew & Wendling, 2010), z něhož plyne, že většina kognitivních schopností podstatných pro porozumění akademickému úspěchu jsou právě úzké schopnosti CHC. McGrew (2012) později rozšířil tento závěr a poukázal na to, že možná nejdůležitější vlastností měřících nástrojů s ohledem na vysvětlení studijního výkonu jsou spíše jejich nároky na zvládnutí kognitivní komplexity než šířka jejich záběru.

Seďm širokých CHC faktorů, které umožňuje měřit WJ IV COG, zahrnují Fluidní inteligenci (Gf), Porozumění – znalosti (Gc), Krátkodobou pracovní paměť (Gwm), Rychlost kognitivního zpracování (Gs), Auditivní zpracování (Ga), Dlouhodobou paměť (Glr) a Vizuální zpracování (Gv) (viz kapitolu 1 a Přílohu A). Dvoutestové úzké klastery jsou k dispozici pro CHC schopnosti Kvantitativního usuzování (RQ), Rozsahu auditivní paměti (MS), Číselných dovedností (N), Rychlosti vnímání (P) a Slovní zásoby (VL). Kognitivní efektivita, která je směsí Rychlosti kognitivního zpracování (Gs) a Krátkodobé pracovní paměti (Gwm) je reprezentována dvou- a čtyřtestovými klastery.

WJ IV OL zahrnuje 12 testů, které rovněž měří schopnosti v širokých oblastech CHC, konkrétně jde o Porozumění-znalosti (Gc), Auditivní zpracování (Ga), Dlouhodobou paměť (Glr) a Krátkodobou pracovní paměť (Gwm). Testy WJ IV OL byly vyvinuty proto, aby zjišťovaly nejdůležitější aspekty spojené s mluveným jazykem, včetně ústního vyjadřování, sluchového porozumění (LS), fonetického kódování (PC), rychlosti přístupu ke slovní zásobě (LA), slovní zásobě samotné⁶⁰ (VL/LD) a rozsahu auditivní paměti (MS).

WJ IV ACH obsahuje 20 testů, které zachycují dvě dříve popsané kognitivní schopnosti, o kterých je pojednáno v Příloze A – Kvantitativní znalosti (Gq) a Schopnosti čtení a psaní (Grw). WJ IV ACH rovněž zahrnuje doplňkové míry Znalostí – porozumění (Gc), Dlouhodobé paměti (Glr)

60 Vyžaduje kombinaci testů WJ IV COG a WJ IV OL.

a Auditivního zpracování (*Ga*). 20 testů WJ IV ACH bylo vyvinuto za účelem měření hlavních aspektů akademického výkonu, včetně čtení, matematiky, psaného jazyka a dalších znalostí kurikula. Specifikace obsahu položek těchto testů byla založena především na cíli poskytnout široký vzorek akademických dovedností, spíše než zprostředkovat hloubkovou diagnostiku v této relativně úzké oblasti.

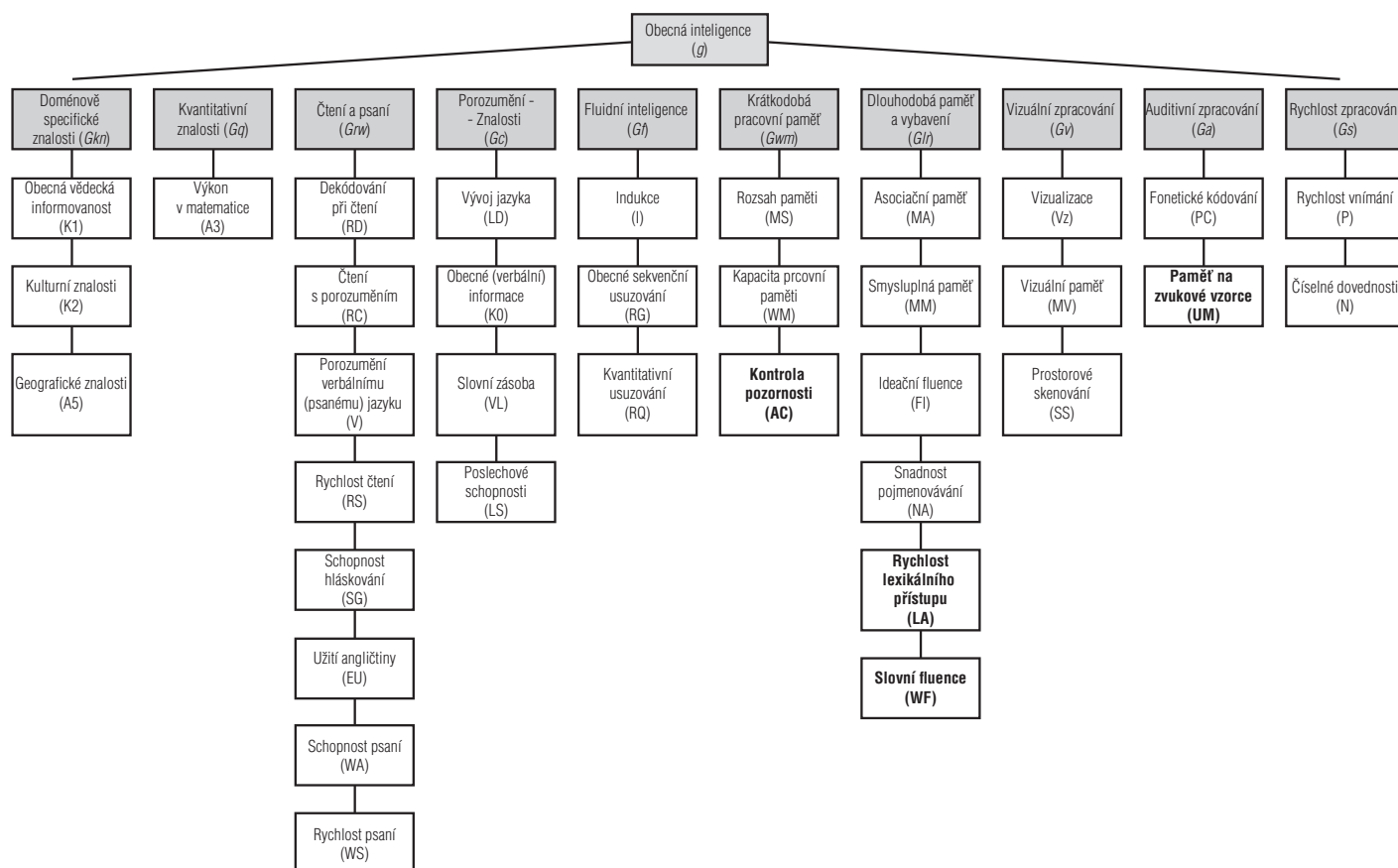
Pokrytí konstruktů, procesů a obsahu

Obrázek 5-1 prezentuje vzájemnou korespondenci širokých a úzkých schopností CHC měřené kompletním WJ IV.

Tato mapa WJ IV ukotvená v poznacích CHC byla vytvořena autory WJ IV. Všechny široké CHC schopnosti jsou reprezentovány nejméně jednou úzkou CHC schopností; většina z nich je nicméně reprezentována dvěma až sedmi úzkými schopnostmi. U většiny testů z původní baterie WJ III, které byly ve WJ IV zachovány, byla jejich implementace do CHC podpořena expertní shodou založenou na vícečetném přezkoumání v rámci tzv. „cross-battery testing“ přístupu (Flanagan, Ortiz, & Alfonso, 2007; Flanagan, Ortiz, Alfonso, & Mascolo, 2006; McGrew & Flanagan, 1998). Přehled společně se srovnáním autora WJ IV a klasifikace nezávislých zdrojů jsou uvedeny v Tabulce 5-1 americké verze manuálu (McGrew, LaForte, Schrank, 2014).

Obrázek 5-1.

*Současné obsahové
vymezení teorie CHC
vzhledem k pokrytí širokých
a úzkých schopností
měřených testy WJ IV*



Pozn.: Tučný font indikuje úzké schopnosti, které reflektují modifikované definice teorie CHC nebo navržené schopnosti založené na výsledcích prezentovaných v tomto manuálu (viz kap. 1 a Příloha A). Materiál v tomto obrázku je adaptován ze Schneider, McGrew (2012).

Tabulka 5-1.

Porovnání autorských
a nezávislých klasifikací
testů WJ III ponechaných ve
WJ IV a autorské klasifikace
nových testů WJ IV

Název testu	Klasifikace autorů WJ IV		Nezávislá klasifikace	
	Široká	Úzká	Široká	Úzká
1. Slovník	Gc	VL/LD	Gc	VL
2. Číselné řady	Gf	RQ, I	Gf	RQ
3. Verbální pozornost	Gwm	WM, Ac	-	-
4. Hledání písmen	Gs	P	-	-
5. Fonologické zpracování	Ga/Glr	Pc/La, FW	-	-
6. Reprodukce příběhů	Glr/Gc	MM/LS	Glr	MM
7. Vizualizace	Gv	Vz	Gv	Vz
8. Všeobecné znalosti	Gc	KO	Gc	KO
9. Formování konceptů	Gf	I	Gf	I
10. Obrácené číselné řady	Gwm	WM/AC	Gwm	WM
11. Hledání čísel	Gs	P	Gs	P
12. Opakování pseudoslov	Ga/Gwm	PC, UM/MS	--	--
13. Audiovizuální učení	Glr	MA	Glr	MA
14. Rozpoznávání obrázků	Gv	MV	Gv	MV
15. Analýza-Syntéza	Gf	RG	Gf	RG
16. Řazení názvů a čísel	Gwm	WM	Gwm	WM
17. Vyhledávání dvojic	Gs/Gwm/Gv	P/AC/SS	Gs	P
18. Paměť na slova	Gwm	MS	Gwm	MS

poskytuje širší popis širokých a úzkých konstruktů měřených WJ IV COG testy, stejně jako charakteristiky podnětových materiálů a žádoucích odpovědí, nároků na úspěšné splnění úkolů a kognitivní procesy, na něž z testů usuzujeme.

Tabulka 5-2.

Obsah testů WJ IV
COG a přehled procesů
a konstruktů

Kognitivní test	Primární široká CHC schopnost Úzká schopnost	Stimul	Požadavky	Kognitivní proces	Odpověď
1. Slovník A. Synonyma B. Antonyma	Porozumění-znalosti (Gc) Znalost lexika (VL) Jazykový vývoj (LD)	Auditorní (slova)	Poslech slova a následné uvedení synonyma/ antonyma	Sémantická aktivace, přístup a přiřazení	Ústní (slova)
2. Číselné řady	Fluidní inteligence (Gf) Kvantitativní usuzování (RQ) Indukce (I)	Vizuální (numerický)	Stanovení číselné sekvence	Reprezentace a manipulace s body na myšlené číselné ose; identifikace a aplikace daného principu/pravidla k dokončení číselné řady	Ústní (čísla)
3. Verbální pozornost	Krátkodobá pracovní paměť (Gwm) Kapacita pracovní paměti (WM) Kontrola pozornosti (AC)	Auditorní (slova, čísla)	Poslech smíšené série čísel a názvů zvířat a zodpovězení specifické otázky o řadě	Kontrolované exekutivní funkce; kapacita pracovní paměti; rekódování akustických verbalizovaných podnětů podržných v bezprostředním vědomí; selektivní zvuková pozornost; kontrola pozornosti	Ústní (slova)
4. Hledání písmen	Rychlost zpracování (Gs) Percepční rychlost (P)	Vizuální (písmena)	Rychlá lokalizace a zakroužkování identických písmen či jejich sledu	Časově omezená vizuální percepce a přiřazování; vizuální diskriminace; ortografické zpracování; dělba pozornosti	Motorická (kroužkování)

Tabulka 5-2. (pokrač.)

Obsah testů WJ IV

COG a přehled procesů

a konstruktů

Kognitivní test	Primární široká CHC schopnost <i>Úzká schopnost</i>	Stimul	Požadavky	Kognitivní proces	Odpověď
5. Fonologické zpracování A. Přístup ke slovům B. Slovní fluence C. Substituce	Auditorní zpracování (Ga) <i>Fonetické kódování (PC)</i> <i>Slovní fluence (GLR-FW)</i> <i>Rychlost přístupu k lexiku (GLR-LA)</i>	Auditorní (slova)	Produkce slova se specifickou fonetickou vlastností; vyjmenování co nejvíce slov začínajících specifickou hláskou; náhrada části slova pro vytvoření slova nového	Sémantická aktivace a přístup, rychlost přístupu k lexiku	Ústní (fráze, věty)
6. Reprodukce příběhů	Dlouhodobá paměť (Glr) <i>Smysluplná paměť (MM)</i> <i>Poslechové schopnosti (Gs-LS)</i>	Auditorní (text)	Poslech a vybavení si detailů z příběhu	Konstrukce propozičních reprezentací a rekódování	Ústní (pasáže příběhu)
7. Vizualizace A. Prostorové vztahy B. Blokované rotace	Vizuální zpracování (Gv) <i>Vizualizace (Vz)</i>	Vizuální (kresby, náčrtky)	Identifikace dvourozměrných prvků, které tvoří tvar; identifikace prostorových rotovaných blokových uspořádání, které odpovídají předloze	Detekce prostorového uspořádání; manipulace (mentální rotace) vizuálních objektů v prostoru; přiřazování	Ústní (písmena) nebo Motorická (ukazování)
8. Všeobecné znalosti A. Kde B. Co	Porozumění-Znalosti (Gc) <i>Obecná (verbální) informovanost (KO)</i>	Auditorní (otázky)	Identifikace toho, kde se typicky nachází určitý objekt a co s tímto objektem lidé obvykle dělají.	Sémantická aktivace a přístup k deklarativním obecným znalostem	Ústní (fráze, věty)
9. Formování konceptů	Fluidní inteligence (Gf) <i>Indukce (I)</i>	Vizuální (kresby)	Identifikace, kategorizace a stanovení pravidel	Na pravidlech založená kategorizace; přepínání mezi pravidly; indukce/vyvozování	Ústní (slova)
10. Obrácené číselné řady	Krátkodobá pracovní paměť (Gwm) <i>Kapacita pracovní paměti (WM)</i> <i>Kontrola pozornosti (AC)</i>	Auditorní (čísla)	Poslech a vybavení si číselné řady v obráceném pořadí	Rozsah chápání a rekódování v pracovní paměti; kapacita pracovní paměti; kapacita pozornosti	Ústní (slova)
11. Hledání čísel	Rychlost zpracování (Gs) <i>Percepční rychlost (P)</i>	Vizuální (čísla)	Rychlá lokalizace a zakroužkování identických čísel ze zadané sady	Časovaná vizuální percepce a přiřazování; vizuální diskriminace; dělba pozornosti	Motorická (kroužkování)
12. Opakování pseudoslov	Auditorní zpracování (Ga) <i>Fonetické kódování (PC)</i> <i>Paměť na zvukové vzorce (UM)</i> <i>Rozsah paměti (Gwm-MS)</i>	Auditorní (bezvýznamová slova)	Poslech bezsmyslného slova a jeho doslovná reprodukce	Analýza řady akustických fonologických prvků v bezprostředním vědomí; efektivita fonologické smyčky	Ústní (slova)
13. Audiovizuální učení	Dlouhodobá paměť (Glr) <i>Asociační paměť (MA)</i>	Vizuální (rébusy) Auditorní (slova)	Učení se a vybavení si obrazových reprezentací slov	Párové asociační kódování v přímé zacílené pozornosti; ukládání a vybavení.	Ústní (věty)
14. Rozpoznávání obrázků	Vizuální zpracování (Gv) <i>Vizuální paměť (MV)</i>	Vizuální (obrázky)	Rozpoznání sady dříve prezentovaných obrázků obklopených obdobnými distraktory	Tvorba ikonických vzpomínek a přiřazování vizuálních stimulů k uloženým vizuálním reprezentacím.	Ústní (slova) nebo Motorická (ukazování)
15. Analýza-Syntéza	Fluidní inteligence (Gf) <i>Obecné sekvenční usuzování (RG)</i>	Vizuální (náčrtky)	Analýza hlavolamů (za použití symbolických formulací) k určení chybějících komponent	Algoritmické uvažování; dedukce	Ústní (slova)

Tabulka 5-2. (pokrač.)
Obsah testů WJ IV
COG a přehled procesů
a konstruktů

Kognitivní test	Primární široká CHC schopnost <i>Úzká schopnost</i>	Stimul	Požadavky	Kognitivní proces	Odpověď
16. Řazení názvů a čísel	Krátkodobá pracovní paměť (Gwm) <i>Kapacita pracovní paměti (WM)</i>	Auditorní (slova, čísla)	Poslech série čísel promíchaných a slov a jejich vybavení ve dvou jinak reorganizovaných řadách	Rekódování akustických verbalizovaných stimulů držených v bezprostřední pozornosti; kapacita pracovní paměti	Ústní (slova, čísla)
17. Vyhledávání dvojic	Rychlost zpracování (Gs) <i>Percepční rychlost (P)</i> <i>Prostorové skenování (Gv-SS)</i> <i>Kontrola pozornosti (Gwm- AC)</i>	Vizuální (kresby)	Rychlá lokalizace a označení opakovaného vzorce	Exekutivní zpracování; kontrola pozornosti; kontrola inhibice a interference; udržení pozornosti	Motorická (kroužkování)
18. Paměť na slova	Krátkodobá pracovní paměť (Gwm) <i>Rozsah paměti (MS)</i>	Auditorní (slova)	Poslech a zopakování sekvence nepříbuzných slov	Tvorba zvukových vzpomínek a verbalizovatelný rozsah zvukového úložiště	Ústní (slova)

Empirické zhodnocení charakteristik obsahu testů

Většina důkazů obsahové validity „obvykle spočívá v konsensuálním expertním posouzení vhodnosti obsahu položek vzhledem ke specifikované oblasti a reprezentativnosti, s níž obsah testu pokrývá obsah této domény“ (Messick, 1989, s. 36). Moderní výzkumy ukazují, že v rámci domén postihujících akademické kurikulum existuje potenciál k posuzování obsahové validity empirickým způsobem. Li a Sireci (2013) ukázali potenciální hodnotu použití multidimenzionálního škálování (angl. *multidimensional scaling*, MDS), podpořeného shlukovou analýzou a korelační analýzou, k hodnocení korepondence mezi empiricky identifikovanými obsahovými dimensemi a profesionálním úsudkem.

MDS je metodou, která představuje doplněk či alternativu k faktorové analýze při práci s velkými multidimenzionálními soubory dat. Jak uvádí APA psychologický slovník (VandenBos, 2007), MDS je „metoda škálování, která reprezentuje vnímané podobnosti mezi podněty tak, že podobné podněty umísťuje do vzájemné prostorové blízkosti, zatímco rozdílné podněty jsou umístěny daleko od sebe“ (s. 599). Tato statistická metoda napomáhá uživatelům porozumět vztahům mezi proměnnými pomocí vizuo-prostorových „map“. Vstupem analýzy je korelační matice sady proměnných; výstupem je vizuo-prostorový graf či mapa, v níž jsou vztahy mezi proměnnými reprezentovány jednak vzdálenostmi, a jednak dimenzemi mapy. Tyto mapy pomáhají uživatelům porozumět klíčovým dimenzím mezi proměnnými. V porovnání s faktorovou analýzou je interpretace MDS spíše kvalitativní a subjektivní. Přesto má však MDS v kontextu obsahové validity unikátní potenciál, protože poskytuje informace „nejen o obsahu, ale i procesech stojících za výkony v různých kognitivních úkolech“ (McGrew, 2005, s. 172).

Multidimenzionální škálování bylo aplikováno i na datové soubory již dříve analyzované explorační či konfirmační faktorovou analýzou, což vedlo „k novému náhledu na charakteristiky testů i konstruktů dříve zakryté statistickou masinérií faktorové analýzy“ (Schneider & McGrew, 2012, s. 110)⁶¹. Jako příklad lze zmínit dnes již klasickou dvoudimenzionální MSD analýzu třiceti kognitivních testů (včetně subtestů Wechslerových inteligenčních škál [WAIS; Wechsler (1955)]) provedenou Marshalkem, Lohmanem a Snowem (1983), Cohenovu, Fiorellovu a Farleyho (2006) trojdimenzionální cylindrickou MDS interpretaci subtestů Wechslerových inteligenčních škál pro děti – čtvrté edice (WISC-IV; Wechsler, 2003), McGrewovo (2012) dvojdimenzionální MDS klastrových skóre kognitivních schopností a školních dovedností WJ-R, a sofistikovanou faktorovou a MDS analýzu 16 kognitivních proměnných ze spojených 38 studií na celkovém vzorku čítajícím přes 8000 respondentů, kterou provedli Tucker-Drob a Salthouse (2009).

61 Pro skvělý popisný a aplikační přehled metod MDS doporučujeme např. Cohen et al. (2006); Süß & Beauducel (2005); Tucker-Drob & Salthouse (2009); a Wilhelm (2005).

Jak jsme uvedli v kapitole 1, jedním z cílů WJ IV je příspěvek disciplíně aplikovaného vývoje testů prostřednictvím implementace současných inovativních modelů sběru a zpracování dat. Následuje stručný popis práce s metodou MDS škálování použitou při vývoji WJ IV jako empirickým nástrojem doplňujícím autorské a expertní posouzení obsahové validity.

Korelace mezi 51 standardizačními a výzkumnými testy WJ IV byly analyzovány dvoudimenzi-
onální MDS procedurou, tzv. *Guttmanovým radexem*; a to v každé z pěti dílčích věkových subpo-
pulací⁶². Radexový model inteligence postulují, že kognitivní schopnosti mohou být organizo-
vány v komplexním radiálním systému (radexu). Jak je popsáno v APA psychologickém slovníku
(VandenBos, 2007):

Radex obsahuje dvě části: (a) simplex, který je relativní vzdáleností od středu kružnice, se schopnostmi, které, čím blíže ke středu kružnice leží, tím blíže jsou i ke konstruktům obecné inteligence, ležícím ve středu samotnému; a (b) circumplexu, který je relativní vzdáleností kolem kružnice, se schopnostmi, které jsou umístěny tím blíže sobě navzájem, čím silněji jsou korelované. Systém tedy umožňuje schopnosti identifikovat v soustavě polárních souřadnic (spíše než v kartézské soustavě užívané jinými systémy). (s. 765)

Tato kapitola obsahuje podrobné výsledky pro věkovou skupinu 9–13 let; kompletní výsledky MDS analýz pro čtyři zbývající věkové skupiny (3–5, 6–8, 14–19, 20–39 a 40–90+) jsou uvedeny v Příloze 1.

Abychom lépe porozuměli sdělení, prezentovanému masou informací v Grafu 5-2 v původním americkém manuálu⁶³ (McGrew, LaForte, Schrank, 2014), měli bychom se především snažit porozumět komplexnímu obrazu. Důležitou informací je, že testy, které sdílí společnou dimenzi, jsou si v mapě typicky prostorově blízké (často tvoří kruhové či oválné shluky, nebo jsou do formy obrazce spojeny úsečkami). Testy, které jsou si vzájemně blízké, interpretujeme jako měřící podobné kognitivní operace nebo sdílející společné podněty či obsahové vlastnosti. V našem příkladu jde o pět velkých barevných oválů.

S ohledem na analýzu obsahu testu jsou tři barevné ovály z Grafu 5-2 charakterizovány jako prů-
niky následujících obsahových dimenzí: *auditivně-lingvistické* (primární testy *Ga*, *Gc* a *Gwm*), *figu-
rálně-vizuální* (primárně *Gv*, a vybrané testy *Gf* a *Glr*) a *kvantitativně-numerické* (primárně *Gq*/*Gf-RQ*
a vybrané testy *Gwm*). Přestože tyto tři dimenze připomínají verbální, figurální a numerickou obsa-
hovou fasetu Berlínského modelu struktury Inteligence (BIS⁶⁴; Süß & Beaudacel, 2005), který je zalo-
žen na extenzivním užití multidimenzionálního škálování, výzkum BIS nezahrnuje do svých modelů
auditivní úlohy. Jak je patrné z grafu 5-2, skupina auditivních testů se shlukuje s verbálními testy
a tvoří širší dimenzi mluveného jazyka nebo dimenzi auditivně-lingvistickou. Toto zjištění naznačuje
možnost, že dimenze verbálního obsahu modelu BIS by potřebovala rozšíření, aby reprezentovala
širší dimenzi mluveného jazyka. Alternativně lze tuto auditivně-lingvistickou dimenzi patrnou v sadě
testů WJ-IV rozdělit dle obsahu na oddělené auditivní a verbální domény, a rozšířit tak počet obsa-
hových dimenzí BIS na čtyři (auditivní, verbální, figurální a kvantitativní). Předmětem výzkumu
v BIS také typicky nebývají výkonové testy čtení a psaní. Z toho důvodu může objevení se nové
dimenze čtení a psaní v testech WJ poukazovat na další chybějící dílek modelu BIS – šestou obsa-
hovou dimenzi (čtení a psaní). Konečně shluk testů WJ IV s rychlostní či fluenční složkou je prav-
děpodobně zapříčiněn spíše podobností nutných kognitivních operací (a nikoli obsahu), přestože
testy čtení, psaní a matematiky s rychlostní složkou vykazují značnou blízkost k dimenzi čtení-psaní
a kvantitativně-numerické obsahové dimenzi. Je vysoce pravděpodobné, že jasnější *diferenciace kog-
nitivního obsahu od kognitivních operací* by se objevila, pokud bychom na standardizačních datech WJ
IV provedli třidimenzionální analýzu MDS (viz Cohen et al., 2006; Tucker-Drob & Salthouse, 2009).

Přestože je pouze předběžná, umožňuje širokospektrální MDS analýza skupin testů WJ IV empi-
ricky podpořit existenci čtyř širokých typů sdílených obsahových charakteristik (konkrétně auditiv-

62 Konstrukce těchto věkových podskupin a korelačních matic je objasněna dále v této kapitole. Podrobnější interpretace metod MDS a jejich výsledků je rozebrána dále v této kapitole. Názvy 51 standardizačních a výzkumných WJ IV testů a jejich zkratky popisuje tabulka 5-10.

63 Vzhledem k tomu, že v ČR máme zatím k dispozici pouze baterii WJ IV Testů kognitivních schopností, výsledky MDS českých dat pochopitelně nemohou přinést takto bohaté výsledky.

64 BIS je heuristický model odvozený jednak z faktorové analýzy, a jednak z analýzy faset pomocí MDS, sloužící pro klasifikaci výkonu v různých úkolech. Neměl by být považován za rysový strukturní model inteligence, jakým je například CHC. Guttmanovy radexové modely MDS nicméně vykazují silné paralely s hierarchickými faktorovými modely zahrnujícími stejné proměnné (Süß & Beauducel, 2005; Tucker-Drob & Salthouse, 2009).

ně-lingvistické nebo auditivní a verbální, figurálně-vizuální, kvantitativně-numerické a čtení-psaní). Tabulka 5-5 v původním americkém Technickém manuálu sumarizuje klasifikaci WJ IV testů v pěti ze šesti věkových skupin (viz také Přílohu H), na nichž byly provedeny MDS analýzy.

Mnoho testů (např. Obrázkový slovník, Číselné řady a Fonologické zpracování) vykazuje velmi konzistentní klasifikace z hlediska obsahu přes věkové skupiny. Jiné testy (např. Obrácené číselné řady) jsou s ohledem na klasifikaci napříč věkovými skupinami konzistentní méně, což napovídá, že v povaze konkrétních úkolů těchto testů existují vývojově podmíněné rozdíly, a nebo, což se jeví jako pravděpodobnější, současné analýzy, jejichž výsledky mohou být zašuměny rozptylem kognitivních operací, poskytují pouze nekompletní obrázek obsahových vlastností testů. Těmto zjevným nekonsistencím v obsahové klasifikaci některých testů můžeme lépe porozumět skrze použití třidimenzionálních modelů MDS.

Zajímavé mohou být i informace o tom, jak se testy v jednotlivých bateriích WJ IV COG, WJ IV OL a WJ IV ACH shlukují. Lze je považovat za širší soubor důkazů o obsahové validitě, zejména s ohledem na organizaci WJ IV do tří trojkomponent. Auditivně-lingvistické uskupení v horním levém kvadrantu MDS Obrázku 5-2 zahrnuje všechny WJ IV OL testy bez rychlostní složky. Horní pravý kvadrant ve stejném obrázku je tvořen pouze WJ IV ACH testy z domén *Grw* a *Gq*. Jedinou výjimku představují Číselné řady, které jsou evidentně mírou kognitivní Gf ⁶⁵. Všechny WJ IV COG testy nalezneme buď v horním levém (auditivně-lingvistický), dolním pravém (primárně *Gs* a *Gwm* testy kognitivní efektivity), nebo dolním levém (primárně figurálně-vizuální a *Gwm* testy) kvadrantu Obrázku 5-2. Až na dříve popsanou výjimku v podobě Číselných řad je nutné zmínit, že žádný z testů WJ IV COG se nevyskytuje v horním pravém „školně výkonovém“ kvadrantu.

Současné zde prezentované důkazy o obsahové validitě založené na přístupu MDS podporují širokou obsahovou strukturu klastrů a baterií WJ IV. Tyto analýzy a jejich interpretace jsou nicméně předběžné a vyžadují extenzivnější práci se standardizačním vzorkem WJ IV, konkrétně komplexnější analýzy smíšenými metodami (faktorovou analýzou a trojdimenzionální MDS), jak je popisují Tucker-Drob a Salthouse (2009). Povzbuzujeme výzkumníky k provedení těchto analýz.

Kognitivní komplexita obsahu a operací

Jak jsme popsali v kapitole 1, důležitým cílem plánu revizí WJ IV bylo zvýšení nároků na *kognitivní komplexitu* konkrétních testů WJ IV. K dosažení tohoto cíle jsme zvolili dva přístupy. První z nich, vcelku obvyklý v rámci aplikovaného vývoje testů, zahrnuje design faktorově komplexních testů CHC, neboli testů, které úmyslně zapojují vliv dvou nebo více úzkých schopností. Tento přístup dokládají příkladem Kaufman a Kaufman (2004a) při vývoji Kaufmann Assessment Battery for Children – Second Edition (KABC-II), kde:

*...se autoři nesoustředili na vývoj „čistých“ úkolů pro měření pěti širokých CHC schopností. Na základě teorie by např. úkoly na *Gv* měly vylučovat ty, které zachycují *Gf* a *Gs*, a testy ostatních širokých schopností, jako *Gc* a *Glr*, by měly měřit pouze tyto schopnosti a žádné jiné. V praxi je ale cílem podrobných testů kognitivních schopností, jakým je i KABC-II, praktické zachycení řešení problémů v různých kontextech a za různých podmínek, k čemuž je zejména na vyšších úrovních nutná určitá komplexita (s. 16).*

V rámci tohoto přístupu k vývoji testů záměrně neeliminujeme ani neminimalizujeme konstruktové irelevantní rozptyl (Benson, 1998; Messick, 1995), a přestože testy měřící více než jednu úzkou schopnost v rámci CHC mají obvykle tendenci k nižší validitě na úrovni indikátorů jednotlivých schopností CHC, podporují naopak další důkazy o validitě (typicky např. vyšší prediktivní validitu).

Druhý přístup ke zvyšování komplexity vybraných testů WJ IV zachovává faktorovou čistotu testů nebo klastrů CHC jak nejvíce je to možné, zatímco úmyslně zvyšuje komplexitu nároků na zpracování informací u testů v rámci konkrétní široké či úzké oblasti CHC (McGrew, 2012). Jak popisují Lohman a Lakin (2011), kognitivní komplexity schopností měřených testy lze dosáhnout buď: (a) zvýšením počtu použitých kognitivních procesů; (b) akumulací rozdílů v rychlosti zpracování v jednotlivých kognitivních komponentách, (c) navýšením počtu důležitějších procesních komponent

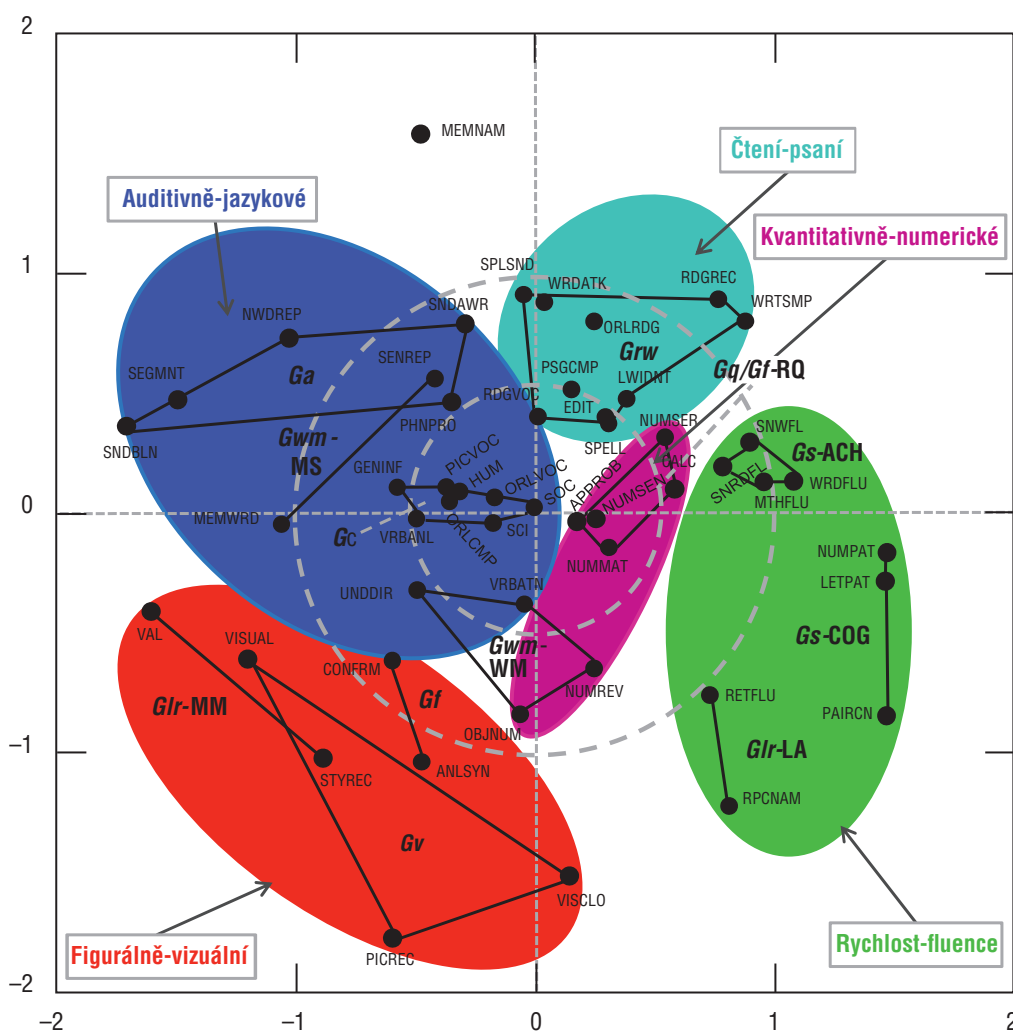
65 Předpovídáme, že v trojrozměrné analýze MDS testů WJ IV by se mělo projevit jasnější oddělení obsahové a kognitivní klasifikace testů. Testy s číselnými řadami mají v testování inteligence dlouhou historii jako silné indikátory kvantitativního usuzování klasifikované pod *Gf* (Carroll, 1993). Podpora pro tuto interpretaci je uvedena v části této kapitoly věnované validitě interní struktury.

(např. úsudků); (d) zvýšením nároků na kontrolu pozornosti a pracovní paměť; nebo (e) navýšením nároků na adaptivní funkce (soustředění, kontrolu a monitoring). Tato druhá forma kognitivní komplexity, kterou je třeba nezaměňovat s faktorovou komplexitou, je definována v kapitole 1 jako takové úkoly, které kladou výraznější nároky na kognitivní zpracování informací (kognitivní zátěž), vyžadují větší alokaci klíčových kognitivních zdrojů (pracovní paměti či kontroly pozornosti), či vyžadují větší zapojení kontrolních či exekutivních funkcí. Díky této formě kognitivní komplexity je cílem vývoje testu zvýšení kognitivní komplexity v rámci konkrétní oblasti CHC, a nikoli úmyslná tvorba smíšených škál zachycujících dvě nebo více schopností CHC.

Pro podporu tvrzení o vysokých nárocích na kognitivní komplexitu testů WJ IV můžeme použít dvě formy důkazů.

Zaprvé, v Guttmanových radexových modelech získaných pomocí MDS považujeme testy blíže středu dvojdimenzionálních grafů za kognitivně komplexnější (Cohen et al., 2006; Marshalek et al., 1983; Tucker-Drob & Salthouse, 2009). (Test s nižší kognitivní komplexitou tedy bude dál od středu grafu MDS). Na Obrázku 5-2 bychom za testy s nejvyšší kognitivní komplexitou považovali ty uvnitř nejmenší kružnice vymezené přerušovanou čarou poblíž středu grafu.

Obrázek 5-2.
Interpretace MDS
(Guttmanův radex)
z hlediska obsahové
validity pro WJ IV ve
věkových skupinách 9
až 13, podsoubor A pro
vývoj modelu americké
standardizace (n = 785)



Pozn.: Tento graf je založen na datech z celé baterie WJ IV, tzn. včetně testů školních dovedností a testů mluveného jazyka. Spíše než k detailnímu studiu je zde uveden proto, aby naznačil způsob, jakým se podobné grafy obvykle interpretují. Testy zobrazené v podobném typu grafu blízko sebe jsou na základě tohoto výsledku považovány za obsahově příbuzné. Skupiny příbuzných testů jsou vybarveny stejnou barvou a souhrnně označeny.

Ty zahrnují většinu Gc testů, testů Aplikovaných problémů (Gq) a Číselných matic (Gf), výzkumných testů Verbálních analogií (Gc/Gf) a Smyslu pro čísla (Gq), dále testů z oblasti čtení a psaní (Grw) – Čteného slovníku, Oprav, Hláskování a Porozumění odstavci, a testu Verbální pozornosti

(Gwm). Příklady testů WJ IV s relativně nízkými nároky na kognitivní komplexitu představují vybrané *Ga* testy (Spojování zvuků, Segmentace) a testy *Gv* (Rozpoznávání obrázků)⁶⁶.

Předběžná analýza MDS testů WJ IV reportovaná v tomto manuálu je vhodná pro přesnou klasifikaci stupně kognitivní komplexity testů WJ IV. Celkem 20 z 51 testů (což je přibližně 40 %) zahrnutých do analýzy MDS reportované grafem 5-2 je součástí baterie WJ IV ACH. Zařazení tak velkého počtu výkonových testů čtení, psaní a matematiky do společné analýzy MDS zkresluje její výsledky směrem k akademickému výkonu. Pro upřesnění míry kognitivní komplexity těchto testů je v budoucnu potřebný další výzkum využívající MDS zahrnující pouze WJ IV COG a WJ IV OL nebo vyváženou sadu indikátorů z každé baterie.

Běžnější akceptovanou metodou analýzy kognitivní komplexity testů je zkoumání faktorového náboje každého z nich na jediném faktoru obecné inteligence (*g*) extrahovaném ze sady testů; což je klasická interpretace připisovaná Arthuru Jensenovi. Jensen (1998) navrhnul, že kognitivní komplexitu testu lze operacionálně kvantifikovat podle jeho náboje na prvním nerotovaném faktoru z faktorové analýzy nebo analýzy hlavních os. Odůvodnění spočívá v úvaze, že testy, které jsou kognitivně komplexnější, vyžadují abstraktní usuzování a řešení problémů a zapojují širší spektrum elementárních kognitivních procesů (Jensen, 1998; Stankov, 2000, 2005), což vyjadřují vysoké náboje na *g*-faktoru. Typicky, ale ne se 100% kongruencí, souhlasí klasifikace testů týkající se příslušnosti k faktoru *g* založené na faktorově-analytickém řešení s výsledky analýzy MDS (Marshalek et al., 1983). Abychom ověřili, zda to platí i v případě testů WJ IV, a to za předpokladu, že náboje na faktoru *g* jsou nejčastěji asociovány s kognitivními testy, s využitím metody faktorové analýzy metodou hlavních os jsme analyzovali 18 testů WJ COG v každé z pěti širších věkových skupin. Náboje každého testu na prvním nerotovaném faktoru v závislosti na věkové skupině uvádí tabulka 5-6 původního amerického manuálu.⁶⁷

Tabulka 5-3.

Náboje WJ IV COG ve faktoru *g* na prvním nerotovaném faktoru získaném metodou hlavních os

Test	Faktor CHC	Věková skupina					Medián
		6–8 <i>n</i> = 823	9–13 <i>n</i> = 1572	14–19 <i>n</i> = 1685	20 – 39 <i>n</i> = 1251	40–90+ <i>n</i> = 1145	
Řazení názvů a čísel	<i>Gwm</i>	0,72	0,69	0,74	0,75	0,79	0,74
Slovník	<i>Gc</i>	0,66	0,67	0,72	0,75	0,75	0,72
Fonologické zpracování	<i>Ga</i>	0,73	0,63	0,68	0,71	0,77	0,71
Formování konceptů	<i>Gf</i>	0,63	0,62	0,66	0,70	0,67	0,66
Obrácené číselné řady	<i>Gwm</i>	0,63	0,59	0,65	0,67	0,69	0,65
Analýza–Syntéza	<i>Gf</i>	0,64	0,63	0,64	0,67	0,70	0,64
Verbální pozornost	<i>Gwm</i>	0,61	0,62	0,64	0,65	0,70	0,64
Číselné řady	<i>Gf</i>	0,62	0,61	0,64	0,59	0,65	0,62
Paměť na slova	<i>Gwm</i>	0,60	0,58	0,61	0,64	0,66	0,61
Vizualizace	<i>Gv</i>	0,60	0,57	0,61	0,66	0,67	0,61
Všeobecné znalosti	<i>Gc</i>	0,44	0,52	0,59	0,63	0,65	0,59
Reprodukce příběhů	<i>Gl</i>	0,58	0,53	0,58	0,54	0,62	0,58
Hledání písmen	<i>Gs</i>	0,57	0,55	0,60	0,57	0,65	0,57
Hledání čísel	<i>Gs</i>	0,56	0,53	0,54	0,53	0,59	0,54
Opakování pseudoslov	<i>Ga</i>	0,55	0,51	0,52	0,51	0,56	0,52
Audiovizuální učení	<i>Gl</i>	0,52	0,47	0,49	0,59	0,60	0,52
Vyhledávání dvojic	<i>Gs</i>	0,45	0,47	0,52	0,51	0,54	0,51
Rozpoznávání obrázků	<i>Gv</i>	0,52	0,43	0,36	0,45	0,47	0,45

Hodnoty nábojů na *g*-faktoru prezentované v tabulce 5-3 lze použít pro hodnocení relativní kogni-

⁶⁶ Je nutné zmínit, že stupeň kognitivní komplexity je relativní vzhledem k souboru testů WJ IV. Jediným způsobem, jak zhodnotit relativní kognitivní komplexitu testů WJ IV vzhledem k jiným bateriím (např. Wechslerovým testům – WPPSI-III, WISC-IV, WAIS-IV [Wechsler, 2002, 2003, 2008], SB-5 [Roid, 2003a], KABC-II [Kaufman, 2004a]) by bylo uskutečnění sběru dat, v jehož rámci by respondencím byly administrovány jak testy WJ-IV, tak i další baterie testů.

⁶⁷ Tuto tabulku zde uvádíme pod číslem 5-3. Podobně jako u klasifikací kognitivní komplexity založených na MDS by náboje ve faktoru *g* v tabulce 5-3 měly být interpretovány pouze jako míra souvislosti s *g*-faktorem v rámci WJ IV. Ke srovnávání napříč různými bateriemi pro testování inteligence je třeba jejich společná analýza (např. s Wechslerovými testy – WPPSI-III, WISC-IV, WAIS-IV [Wechsler, 2002, 2003, 2008], SB-5 [Roid, 2003a], KABC-II [Kaufman, 2004a]).

tivní komplexity každého z testů WJ IV COG. Například Slovník ($Mdn\ g^{68} = 0,72$) je zřejmě nejvíce kognitivně komplexní mírou Gc . Např. Obecné informace (Gc) mají znatelně nižší střední hodnotu g -náboje ($Mdn\ g = 0,59$). V rámci Gv by Vizualizace ($Mdn\ g\ 0,61$) byly klasifikovány jako kognitivně komplexnější než Rozpoznávání obrázků ($Mdn\ g\ 0,45$).

Doména auditivního zpracování (Ga) je obzvláště osvětlující příklad. Jak je popsáno v kapitole 1 Příručky examinátora WJ IV (Mather & Wendling, 2014b), nový test WJ IV Fonologické zpracování byl navržen tak, aby vyžadoval komplexnější kognitivní zpracování většího počtu fonetických kódovacích funkcí (Ga -PC). Výsledky analýz hlavních os napovídají, že Fonologické zpracování je kognitivně mnohem komplexnější ($Mdn\ g\ 0,71$) než např. Opakování pseudoslov ($Mdn\ g = 0,52$). Fonologické zpracování je ve skutečnosti jedním ze tří testů v rámci WJ COG s nejvyšším nábojem na g -faktor. Z výsledků analýz hlavních os je patrné, že cíle vývoje testu, kterým bylo zvýšení kognitivní komplexity vybraných testů WJ IV bez nutného navýšení faktorové komplexity, bylo dosaženo.

Vývojové křivky klastrů schopností WJ IV

Z přehledu souhrnných deskriptivních statistik v Přílohách B a C je patrné, že všechny testy a klastry WJ IV vykazují změny průměrných skóre, které jsou konzistentní s vývojově podmíněným růstem a poklesem kognitivních schopností a školních dovedností napříč celým věkovým spektrem. Rozdíly ve vývojových křivkách podporují tvrzení o existenci rozdílných, unikátních schopností (Carroll, 1993). Grafy 5-3 a 5-4⁶⁹ v původním americkém manuálu prezentují příklady vývojových křivek význačných klastrů WJ IV COG ve věkovém rozpětí od 5 do 80 let. Z vývojových křivek je patrné, že unikátní schopnosti měřené WJ IV sledují rozdílné trajektorie v průběhu vývoje, a to od dětství až do stáří. Ilustrační křivky mají svůj začátek v 6 letech a 0 měsících (6-0). Byly sestaveny odečtením normativního referenčního W -skóre (REF W^{70}) pro věk 6-0 od všech následujících REF- W tohoto klastru až po věkovou skupinu 90+⁷¹. Touto procedurou jsme zkonstruovali vývojové křivky se společným začátkem v nule. Věk 6-0 byl vybrán jako počátek z toho důvodu, že počínaje od této věkové úrovně výš všechny klastrové skóre WJ IV disponují normativním REF W . Vývojové křivky v Grafech 5-3 až 5-8 pro všechny věkové úrovně byly zkonstruovány ze skóre stejných respondentů.

68 $Mdn\ g$ = medián náboje přes věkové podskupiny.

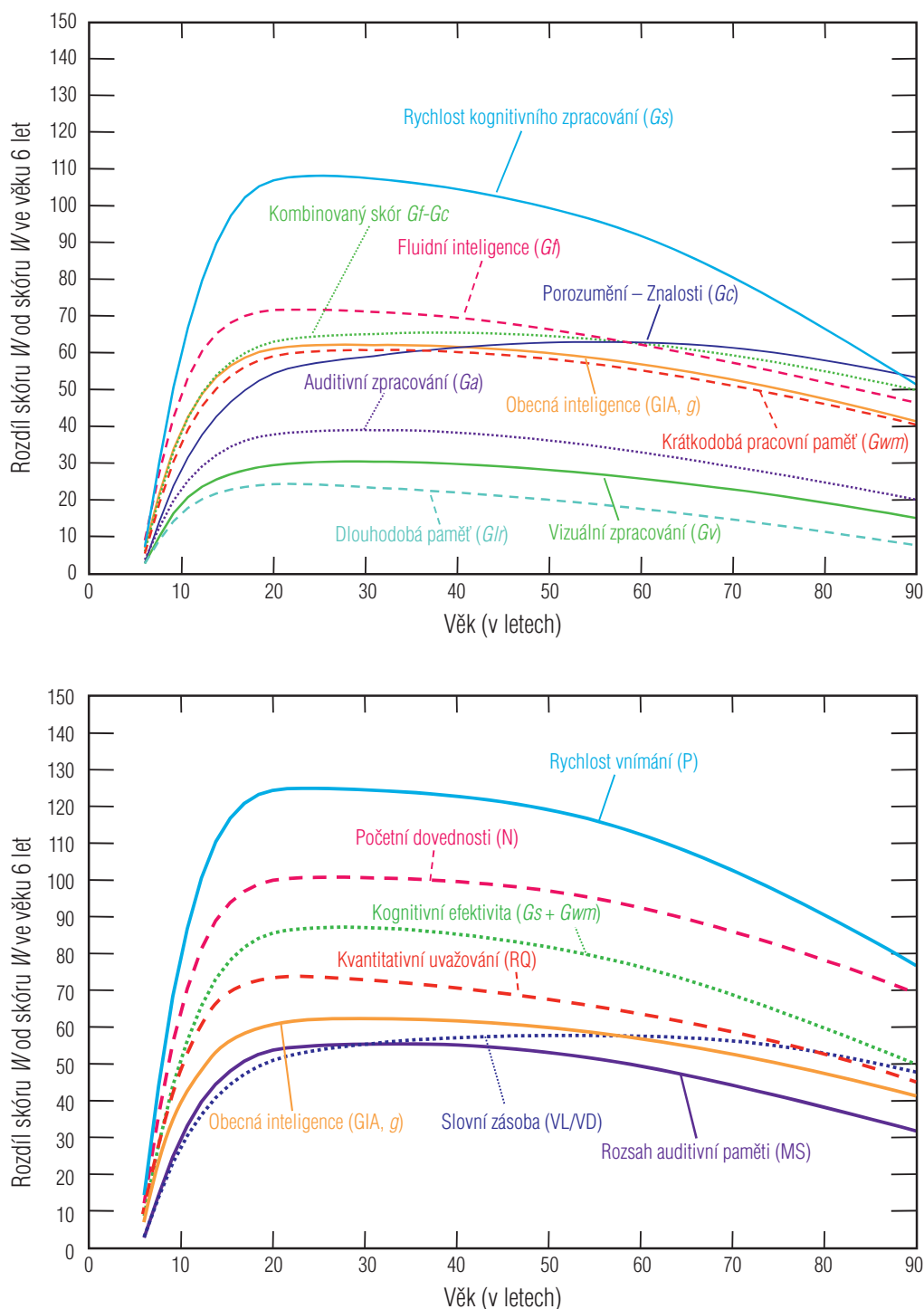
69 Zde uvedeny společně na obrázku 5-3.

70 Referenční W skóre definuje kapitola 3.

71 Pokud existují dvě verze klastru WJ (např. Fluidní usuzování a Fluidní usuzování rozšířené), zde je uveden pouze klaster tvořený dvěma testy.

Obrázek 5-3.

Přetištěné obrázky 5-3 a 5-4 z amerického technického manuálu opatřené českými popisky. Skóry klastrových skóre sedmi faktorů CHC a dalších klastrů s odečtenou referenční hodnotou pro věk 6 let podle věku.



Existence unikátních vývojových vzorců pro většinu širokých i úzkých schopností WJ IV napříč doménami CHC i v jejich rámci je jednou z forem důkazů o tom, že v součinnosti s informacemi o obsahu testů, jejich struktuře a vztahu s dalšími proměnnými (o nichž bude pojednáno dále v této kapitole) jsou skóry WJ IV validní pro měření kognitivních, jazykových i znalostních schopností. V souvislosti s grafy prezentovanými v této kapitole si dovolueme jedno upozornění: přestože grafy popisujeme jako *vývojové křivky*, jedná se o vizualizaci průřezových, nikoli longitudinálních dat. Znázorňují nárůst a pokles v mediánu výkonu v konkrétních věkových skupinách v obecné populaci v době standardizace WJ IV, nikoli vývoj v rámci jedné kohorty respondentů napříč časem.

Obecné (g), široké a úzké shluky WJ IV Cog

Vývojové křivky Obecné intelektové schopnosti (*General Intellectual Ability*; GIA, g), Složeného skóru Gf-Gc a sedmi klastrů kognitivních faktorů⁷² CHC jsou vykresleny v Obrázku 5-3. Vzorce růstu a poklesu sedmi kognitivních faktorů CHC z WJ IV se značně liší, což podporuje náš předpoklad o existenci rozdílných schopností. Faktorový klaster dlouhodobé paměti (*Glr*) podléhá vývojovým změnám méně než jiné kognitivní schopnosti měřené WJ IV COG. Vývojová křivka Dlouhodobé paměti má tvar, který je typický pro testy kognitivních schopností, u nichž je výkon založen na schopnosti, která není tolik ovlivňovaná formálním tréninkem a učením. Vývojová křivka *Glr* (a částečně i křivky *Ga* a *Gv*) je příznačná pro schopnosti, které se vyvíjí spíše v součinnosti s neformálním a nepřímými zkušenostmi učení, které se typicky v průběhu věku mění relativně málo. Naopak vývojová křivka faktorového klastru Porozumění-Znalosti (*Gc*) jsou příkladem měření, kde hraje zkušenost s přímým a formalizovaným učením důležitou roli. U faktoru *Gc* můžeme pozorovat spojitý růst až daleko do dospělosti (věkového rozmezí 55–70); oproti tomu *Glr* dosahuje vrcholu o dost dříve, a to přibližně ve 20 letech. Křivka *Gc* navíc během pozdní dospělosti klesá daleko méně než křivka *Glr*, u níž pozorujeme poměrně strmý pokles v mnohem nižším věku (přibližně 30 a níže). Křivka *Gc* je křivkou kognitivní schopnosti, jejíž skóry jsou napříč vývojovým spektrem daleko konzistentnější než u schopností reprezentovaných křivkami klastrů *Glr*, *Gv* a *Ga*.

Povšimněme si rovněž strmého nárůstu ve faktorovém klasteru Rychlosti zpracování (*Gs*), patrného na Obrázku 5-3. Tato schopnost dosahuje své horní asymptoty mezi 20. a 30. rokem a následně postupně klesá, a to po zbytek celého věkového spektra⁷³. Faktorové klastery Krátkodobé pracovní paměti (*Gwm*) a Fluidního usuzování (*Gf*) dosahují svého stropu rovněž mezi 20 a 30 lety a vykazují podobný pokles po zbytek života.

Jak jsme předpokládali na základě toho, že je GIA (g faktor) tvořen sedmi testy (jeden z každé ze sedmi CHC kognitivních domén), nachází se vývojová křivka tohoto faktorového klastru přibližně uprostřed prostoru vymezeného ostatními křivkami na Obrázku 5-3. Obecná inteligence (g) měřená faktorovým klastrem GIA, dosahuje svého vrcholu mezi dvacítkou a třicítkou. Složený skór Gf-Gc, který rovněž zahrnuje více než jednu širokou doménu schopností CHC, v rané fázi vykazuje podobný růstový trend jako GIA (g), nicméně začíná se odchylovat směrem nahoru přibližně ve 20 letech, a tím vykazuje delší období růstu než GIA. Tato prodloužená růstová perioda v dospělosti odráží zařazení schopností z domény krystalizované inteligence (*Gc*) do kompozitu *Gf-Gc*.

Druhá část Obrázku 5-3 demonstruje existenci rozdílných vzorců růstu a poklesu schopností měřených úzkými faktory schopností WJ IV COG a klinickými klasteru Rozsahu auditivní paměti (*Auditory Memory Span*, MS), Slovníku (*Vocabulary*, VL/LD), Kvantitativního usuzování (*Quantitative Reasoning*, RQ), Kognitivní efektivity (*Gs+Gwm*), Číselných dovedností (*Number Facility*, N) a Percepční rychlosti (*Perceptual Speed*, P). Křivka Obecné kognitivní schopnosti (GIA, g) je v grafu přítomna pro srovnání. Podle očekávání jsou křivky Rozsahu auditivní paměti (MS) a Slovníku (VL/LD) podobné svým odpovídajícím širokým schopnostem *Gwm* a *Gc* v první části Obrázku 5-3. Křivka Kvantitativního usuzování (RQ) rovněž připomíná křivku široké *Gf* z první části Obrázku 5-3. A konečně, vzhledem k zařazení testů s rychlostní složkou do těchto klastrů pozorujeme vyšší maxima a strmější pokles u Kognitivní efektivity (*Gs+Gwm*), Číselných dovedností (N) a Percepční rychlosti (P), což je patrné i u křivky *Gs* v Grafu 5-3.

Vývojové křivky v datech české standardizace

Předchozí pasáž popisovala výsledky americké standardizační studie. Na základě českých dat byly vytvořeny obdobné grafy vývojových křivek (Obrázky 5-4 a 5-5), které umožňují porovnat nárůst a pokles W-skórů mezi 5. rokem věku (na rozdíl od amerických grafů, kde byl počátek stanoven na 6

72 V celé této kapitole jsou používány příbuzné pojmy *klaster* a *faktor*. Je důležité definovat, co každý z nich znamená. Když je použit pojem *klaster*, odkazuje k manifestnímu nebo získanému skóru složeného shluku WJ IV. Když je použit pojem *faktor*, odkazuje k latentnímu konstruktivnímu nebo dimenzi CHC, který má klaster reprezentovat. Slovní spojení *faktorový shluk* odkazuje ke shluku WJ IV (pozorovanému nebo získanému skóru), který představuje latentní faktor nebo schopnost teorie CHC (viz kap. 1 a Přílohu A).

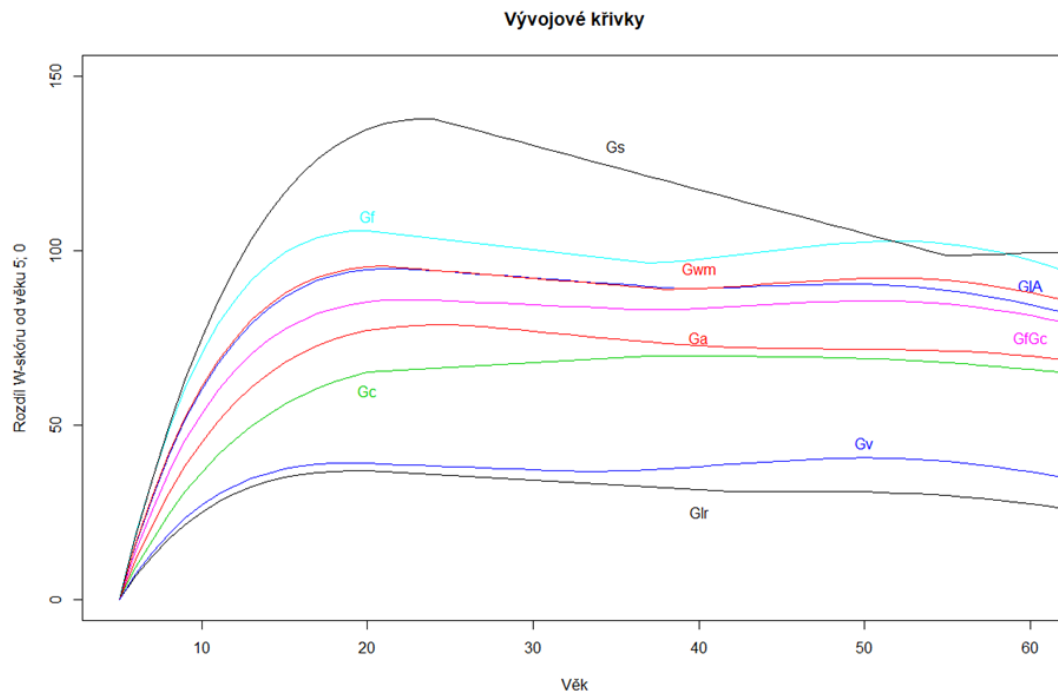
73 Časované testy mají obecně širší rozpětí schopností probandů na škále W a větší směrodatné odchylky než nečasované testy. Tento velký rozsah schopností W se odráží v relativně prudkých růstových křivkách pro shluky obsahující jeden nebo více časovaných testů (jako je *Gs*). Uživatele odkazujeme na kap. 2 „Kalibrace časovaných testů“, kde najdou více informací, jak interpretovat skóry časovaných testů a klastrů obsahujících časované testy. Viz také další diskusi v McGrew, Werder, & Woodcock (1991).

let) a 60. rokem věku. Na osách y jsou vyneseny hodnoty diferenčních skóreů – od každého ze skóreů byly odečteny průměrné hodnoty tohoto skóre ve věku 6; 0.

Náš Obrázek 5-4 lze přímo porovnat s americkým grafem v první části Obrázku 5-3 a náš Obrázek 5-5 s americkým grafem v druhé části Obrázku 5-3. Přestože tvary křivek jsou mírně odlišné, pravděpodobně vlivem podstatně menšího výběrového souboru českých standardizačních dat, jejich umístění je v podstatě ve shodě s umístěním odpovídajících křivek v grafech amerického technického manuálu.

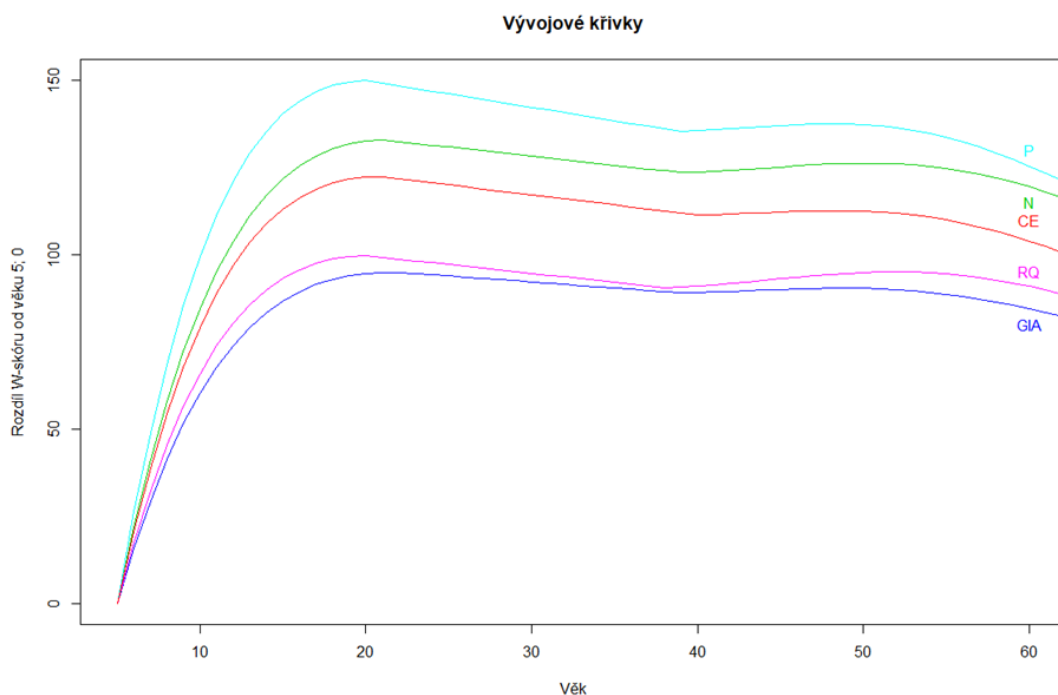
Obrázek 5-4.

Vývoj diferenčních hodnot klastru GIA a dalších sedmi faktorových klastrů a složeného skóre Gf-Gc podle věku



Obrázek 5-5.

Vývoj diferenčních hodnot klastru GIA a čtyř dalších kognitivních schopností podle věku



Vnitřní struktura WJ IV

Primárním zdrojem důkazů o validitě, které jsou relevantní vzhledem k vnitřní struktuře pedagogických a psychologických testů, je míra, do níž testové skóry utvářejí vztahy implikované souvisejícími teoretickými konstrukty (AERA et al., 1999, 2001). Uvádíme dvě formy důkazů o validitě vnitřní struktury WJ IV. Nejprve je popsán vzorec interkorelací mezi skóry klastrů. Následně jsou provedeny explorační a konfirmační multivariační analýzy vztahů mezi testy WJ IV.

Interkorelace testů a klastrů ve standardizačním výběru

Americká verze technického manuálu v Přílohách E a F⁷⁴ uvádí interkorelace mezi všemi testy a klastry WJ IV. Korelační matice v Přílohách E a F jsou k dispozici pro šest širokých věkových kohort (3 až 5, 6 až 8, 9 až 13, 14 až 19, 20 až 39, 40 až 90+). Jak již bylo napsáno v kapitole 3, přijatelné postupy generování dat použité jako součást záměrně nekompletního plánu sběru dat standardizační studii přinesly kompletní data pro všechny testy WJ IV (velikosti výběrových souborů se pohybovaly od 6637 do 7416 a lišily se z důvodů různého minimálního věku vhodného pro administraci některých testů). Testové a klastrové interkorelační matice byly spočítány na těchto velkých, relativně kompletních datech, s uplatněním metody vymazání celého záznamu (tzv. „*listwise deletion*“⁷⁵) v případě chybějících dat. Výsledné korelace testů a klastrů jsou všechny založeny na stejném souboru respondentů. Jak je uvedeno v tabulkách E-2 až E-6 v přílohách, korelační matice pro věkové kohorty 6 let a starší jsou založeny na všech testech WJ IV, s velikostmi vzorku v rozpětí od 825 do 1 685 respondentů. V případě předškoláků (3–5 let) byly korelace spočítány pouze pro 28 subtestů vhodných s ohledem na věk. Velikosti vzorků ve skupině předškoláků napříč testy variovaly a jsou k dispozici v sekci Příloh v tabulce E-1. Velikosti vzorků pro interkorelace klastrů v Příloze F, konkrétně tabulkách F-1 až F-6, jsou v rozmezí od 435 do 1 685 osob.

Vztah a síla korelací mezi skóry testů a klastrů nám může poskytnout důkazy, že vztahy mezi skóry odpovídají teoretickým předpokladům, které o konstruktech máme (AERA et al., 1999, 2001; Campbell & Fiske, 1959). Interkorelace testů a klastrů (viz Přílohy E a F) poskytují empirickou podporu pro několik závěrů o vztazích mezi skóry testů WJ IV. Zaprvé, korelace jsou obecně vyšší mezi doménami a klastry, které považujeme za příbuzné v rámci CHC, než mezi nepříbuznými. Například z Tabulky E-3 (Příloha E, věk 9–13 let) je patrné, že korelace Slovníku (Gc) s dalšími dvěma Gc testy (Ústní porozumění a Obrázkový slovník) se pohybují v rozmezí od 0,62 do 0,70, zatímco korelace se třemi COG testy zaměřených na Gs (Hledání písmen, Hledání čísel a Vyhledávání dvojic) dosahují pouze hodnot od 0,26 do 0,32. Oproti tomu korelace mezi těmito třemi COG Gs testy se pohybují od 0,57 do 0,60. Příklad na úrovni klastrů je k nalezení v Příloze F, Tabulce F-4 (věková kohorta 14–19 let).

Zadruhé, rozsah korelací širokých kognitivních klastrů CHC (typicky od 0,30 po 0,60) je nižší než těch, které pozorujeme mezi primárními klastry školních dovedností. To nám poskytuje další podporu pro tvrzení, že klastry WJ IV COG skutečně měří rozdílné kognitivní schopnosti. Zatřetí, korelace mezi klastry školních dovedností jsou typicky vyšší mezi klastry ze stejné domény školních dovedností, než jaké pozorujeme mezi klastry z odlišných domén. Např. korelace mezi klastrem Základních schopností čtení a dvěma klastry Čtení s porozuměním jsou 0,74 a 0,78. Oproti tomu klaster Základních dovedností čtení koreluje slaběji se skóry čtyř matematických klastrů (konkrétně v rozmezí od 0,56 do 0,62).

Multivariační statistické postupy (např. shluková analýza, multidimenzionální škálování, explorační a konfirmační faktorová analýza) jsou nápomocné při rozkrývání vzorců konvergentních i divergentních souvislostí testů tvořících korelační matice v Příloze E. Výsledky těchto procedur popíšeme dále v této kapitole.

74 V této české verzi Technického manuálu jsou výsledky analogicky zaměřených analýz uvedeny dále v kapitole „Doklady strukturní validity založené na analýzách dat české standardizace“.

75 Pozn. překl.: Metoda práce s chybějícími daty, kdy je při absenci hodnoty jedné proměnné odstraněn celý řádek v datové matici.

Důkazy o vnitřní struktuře: Trojstupňová analýza strukturní validity WJ IV

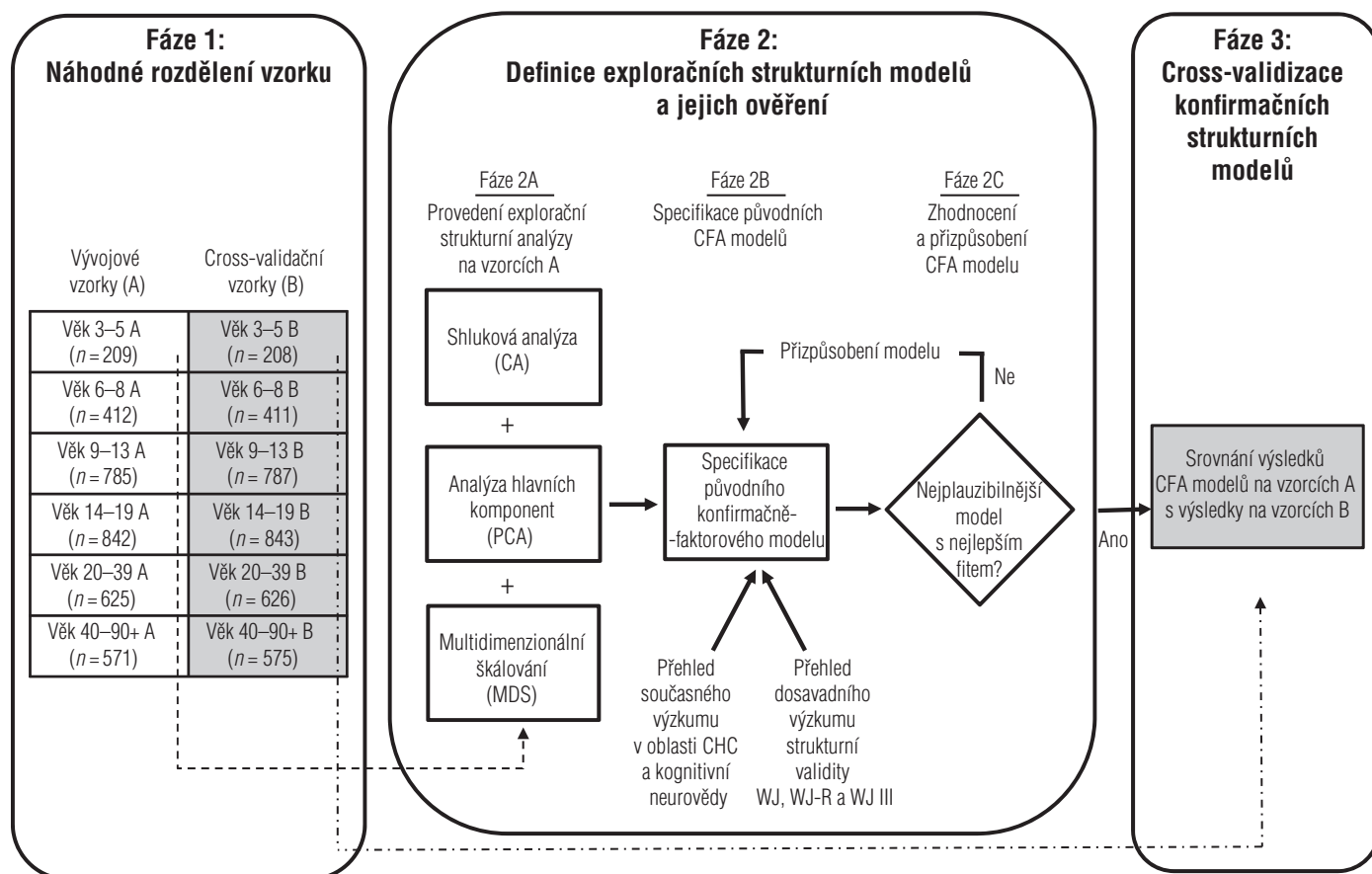
Primárním indikátorem vnitřní, nebo také strukturní validity pedagogických a psychologických testů je míra, „ve které vzájemné vztahy mezi položkami nebo částmi testu odpovídají konstrukt, na kterém je navrhovaná interpretace testového skóru založena“ (AERA et al, 1999, 2001, s. 20). Výhodou strukturních analýz WJ IV jsou jejich široké základy, které leží v komplexních strukturních analýzách WJ (Woodcock, 1978), WJ-R (McGrew, Werder, & Woodcock, 1991) a WJ III (McGrew & Woodcock, 2001). Braden a Niebling (2012) zhodnotili kvalitu prováděných důkazů o validitě WJ III, na kterých je založen i WJ IV, jako silnou, a to přidělením nejvyššího možného celkového hodnocení 5 bodů na 5-bodové škále *kvality důkazů o validitě*. V přehledové studii 20 let výzkumu v oblasti schopností CHC měřených různými intelektovými bateriemi konstatovali Keith a Reynolds (2010), že CHC strukturu WJ-R a WJ III podporuje rozsáhlá řada výzkumných zjištění.

Během raných fází sběru dat WJ IV byla provedena série exploračních shlukových a faktorových analýz a rovněž konfirmačních faktorových analýz. Tyto analýzy byly provedeny napříč celým věkovým rozsahem pro prvních 1 517 účastníků standardizační studie. Primárním cílem těchto analýz bylo určit, jak jsou nové nebo revidované testy syceny různými kognitivními, jazykovými či školně-výkonnými faktory. Tento krok, který zahrnoval rovněž logické a psychologické obsahové analýzy testů, umožnil modifikaci či eliminaci určitého malého počtu testů. V tomto bodě byly zformovány vztahy mezi testy WJ IV a širokými faktory CHC a předběžná organizační struktura WJ IV.

Po ukončení sběru standardizačních dat (N = 7416) byl k ověření vnitřní strukturní validity baterie WJ IV uplatněn trojstupňový proces znázorněný Obrázkem 5-6.

Obrázek 5-6.

Trojstupňová procedura
ověřování strukturní validity
WJ IV



Fáze 1: Náhodné rozdělení vzorku

Jak je zobrazeno na Obrázku 5-6, standardizační výběr WJ IV byl rozdělen na šest skupin v závislosti na věku (3-5; 6-8; 9-13; 14-19; 20-39 a 40-90+). Každá z těchto skupin byla následně náhodně rozdělena na výběr pro vývoj modelu (MD; vzorek A) a cross-validační výběr (MCV; vzorek B) o přibližně stejné velikosti (viz Fáze 1 na Obrázku 5-6). Výsledných 12 výběrů obsahovalo od 208 do 834 účastníků standardizační studie WJ IV.

Jak je popsáno v kapitole 1, během vývoje WJ IV bylo několik testů WJ III vynecháno a několik nových testů bylo přidáno. Některé z testů WJ III byly přepracovány, aby byly více kognitivně-komplexní (viz kapitola 1 obsahující definice). Navíc, současná teorie CHC, podpořená neurokognitivním výzkumem, významně ovlivnila vývoj a revize WJ IV. Vzhledem k různým charakteristikám změn a jejich velkému počtu jsme usoudili, že striktně konfirmační přístup vycházející z CHC, který byl použit v případě WJ III, by mohl být nedostatečný pro identifikaci nových strukturních dimenzí a vztahů. Namísto toho jsme při validizační strategii postupovali systematickým exploračním způsobem, vedoucím k vygenerování modelů a jejich validizaci na nezávislých souborech dat.

Přístup k WJ IV zahrnující rozdělení vzorku a vícestupňové exploračně-konfirmační řešení považujeme za nejvědeckější způsob prozkoumání strukturní validity jakékoli současné baterie kognitivních, jazykových a školně-výkonových testů. Důkazy o strukturní validitě pro většinu kognitivních a školně-dovednostních baterií v posledních letech byly shromažďovány primárně s pomocí konfirmační faktorové analýzy (CFA) (viz Keith & Reynolds, 2012; excelentní popis využití metod CFA při validizaci testů inteligence). Baterie WJ III se spoléhala výlučně na CFA. Stejně tak většina dalších kognitivních baterií (DAS-II [Elliot, 2007]; KABC-II [Kaufmann & Kaufmann, 2001a]; SB-F [Roid, 2003a]; WAIS-IV [Wechsler, 2008]) k ověření strukturní validity využívala přístup CFA. Oproti tomu WISC-IV (Wechsler, 2003) kombinoval explorační faktorovou analýzu (EFA) a metody CFA.

Analýza strukturní validity ve všech šesti věkových kohortách

Výběrový soubor od 9 do 13 let určený pro vývoj modelu (MD) byl úmyslně zvolen jako první, který měl být analyzován zmíněným trojstupňovým procesem ověřování strukturní validity. Tato věková skupina představuje jednu z hlavních věkových skupin, u kterých je baterie WJ IV typicky používána. Dokončení trojstupňového procesu popsaného Obrázkem 5-6 ve věkové skupině od 9 do 13 let a použití výsledků z této věkové skupiny pro definici přibližného počátečního modelu (Fáze 2B) pro všechny ostatní věkové skupiny umožnilo větší efektivitu komplexního trojstupňového vícevýběrového procesu strukturní validizace. Ještě důležitější je, že tato strategie je založena na předpokladu vycházejícím z inspekce výsledků explorační studie přes všechny věkové skupiny (Fáze 2A na Obrázku 5-6), že v rámci baterie WJ IV existuje společná strukturní organizace napříč všemi věkovými skupinami. Jakékoli rozdíly mezi výsledky věkových skupin se nemanifestují přítomností nebo absencí širokých nebo úzkých faktorů, ale místo toho reflektují věkové rozdíly ve faktorových nábojích specifických testů a reziduálních korelací testů. Výjimku ze společného přibližného modelu představuje skupina dětí v předškolním věku (věk 3 až 5 let). Model pro tuto skupinu obsahoval menší skupinu testů (28) než soubory ostatních věkových skupin (kterým byla administrována sada 51 testů).

Vzhledem k významné roli, kterou věková skupina 9-13 hrála v poskytnutí předběžného výchozího bodu pro stejné procedury ve všech dalších vzorcích, jsou všechny výsledky z této věkové skupiny popsány a interpretovány v této kapitole. Výsledky trojstupňového ověření strukturní validity ve všech věkových skupinách jsou popsány v tabulkách, které následují v této kapitole.

Explorační faktorová analýza (EFA) je nejvhodnějším nástrojem ve chvíli, kdy je testová baterie nejprve zkonstruována, a pak následně revidována a reorganizována, nebo v případě, kdy během vývoje testu nemáme k dispozici dostatečně silnou teorii, která by design testu směřovala. Metody konfirmační faktorové analýzy (CFA) oproti tomu preferujeme ve chvíli, kdy máme o vnitřní struktuře nástroje dostatečné povědomí, ať už díky vícenásobným revizím konkrétního testu, nebo v případě, že existuje silná teorie konstruktů, které test zjišťuje (Brown, 2006; Kline, 2011; Schreiber, Nora, Stage, Barlow, & King, 2006).

U CFA založené na strukturním modelování Jöreskog (1993) rozlišuje mezi metodami striktně konfirmačními (*strictly confirmatory*, SC), analýzami alternativních modelů (*alternative models*, AM)

a generováním modelů (*model generating*, MG)⁷⁶. Striktně konfirmační metody (SC) formulují a testují jeden model, zatímco při analýze AM výzkumník specifikuje více možných „konkurenčních“ nebo alternativních modelů, z nichž se snaží vybrat model (modely) s nejlepší shodou s daty. Podle Jöreskoga (1993), „je zdaleka nejběžnější situací generování modelů (MG)“ (s. 295). Přístup MG je charakterizován (a) specifikací počátečního modelu založeného na věcné teorii, výzkumných hypotézách nebo předběžných nápadech; (b) odhadem počátečního modelu a hodnocením jeho celkového fitu a dalších statistik (např. standardizovaných reziduí a modifikačních indexů), které naznačují možné úpravy nebo „ladění“ modelu; (c) specifikací a odhadem revidovaného modelu; (d) iterací kroků (b) a (c) tak dlouho, dokud není nalezen model, který „se rozumně shoduje s daty vzorku a v němž jsou všechny parametry smysluplné a věcně interpretovatelné“ (Jöreskog, 1993, s. 313).

Přístup MG je obzvláště vhodný pro vývoj aplikovaných psychologických testů. Přístup MG představuje hlavní komponentu ve víceúrovňovém ověřování strukturní validity WJ IV popsáném touto kapitolou. Nicméně přístup MG „vydělává na náhodě“ (angl. *capitalizes on chance*), neboli představuje jistou formu „rybaření v datech“ a vede ke zkreslení inferenčních koeficientů shody. Vzato do extrému, Wagenmakers, Wetzels, Borsboom, van der Madd, & Kievit (2012) varují:

„obracíte svůj soubor dat ze všech stran, pátráte po zajímavých vzorech, použili jste data, aby vám pomohla formulovat specifickou hypotézu. Přestože data mohou po dokončení těchto expedicí stále sloužit mnoha účelům, existuje jeden účel, pro který už nejsou vhodná – a to konkrétně pro testování hypotéz, jejichž formulaci napomohla“ (s. 633)

V obraně před tím, co Wagenmakers a jeho kolegové (2012) popisují jako „bagrování“ nebo dokonce „mučení“ dat, nicméně umožňuje získat vzhled z exploračních vědeckých metod vedoucím k novým objevům a pokroku, doporučuje několik výzkumných autorit validizaci finálních MG modelů na nezávislých souborech dat (Brown, 2006; Jöreskog, 1993; Wagenmakers et al., 2012). Rozdělení standardizačního souboru WJ IV na podsoubory pro vývoj modelu (MD) a jeho validizaci (MCV) nám dovolila zavést postupy, která jsou ve shodě s doporučeními té nejlepší praxe při kombinování exploračních metod MG a CFA.

Fáze 2: Explorační generování strukturních modelů (MG) a jejich ověření na souborech pro vývoj modelu

Každý ze šesti souborů pro vývoj modelu byl analyzován třemi různými exploračními multivariačními metodami: shlukovou analýzou (CA), explorační analýzou hlavních komponent (PCA) a multidimenzionálním škálováním (viz Fáze 2a Na Obrázku 5-6). Použití těchto tří metodologických optik umožňuje detailní prozkoumání vztahů mezi kompletní sadou testů WJ IV. Tento přístup v duchu „nenechat kámen na kameni“ přináší podrobnou explorační strukturu baterie WJ IV nezávislou na apriorním pohledu (např. striktním modelu CHC)⁷⁷. Přestože zmíněné tři optiky analýzy dat jsou zde definovány a ilustrovány na vzorku věkové skupiny 9-13 let určeném pro generování modelu, podobné sady exploračních analýz pro všechny ostatní věkové skupiny jsou uvedeny v Přílohách G (CA), H (PCA) a I (MDS).

Fáze 2A: Explorační shluková analýza

Výsledky shlukové analýzy (CA) pro věkovou skupinu 9-13 let jsou prezentovány na obrázku 5-10 v původním americkém manuálu⁷⁸. Výsledky shlukové analýzy pro další věkové kohorty jsou k nahlédnutí v Příloze H. Podle Aldenderfer & Blashfielda (1984) je shluková analýza

„generickým označením pro širokou škálu postupů, které mohou být použity pro vytvoření klasifikace. Tyto postupy empiricky vytváří „shluky“ nebo skupiny velmi podobných entit. Specifičtěji vzato je metoda shlukování multivariační statistickou procedurou, jejímž vstupem je soubor dat obsahující informace o výběrovém souboru entit, které se mají reorganizovat do relativně homogenních skupin“ (s. 7)

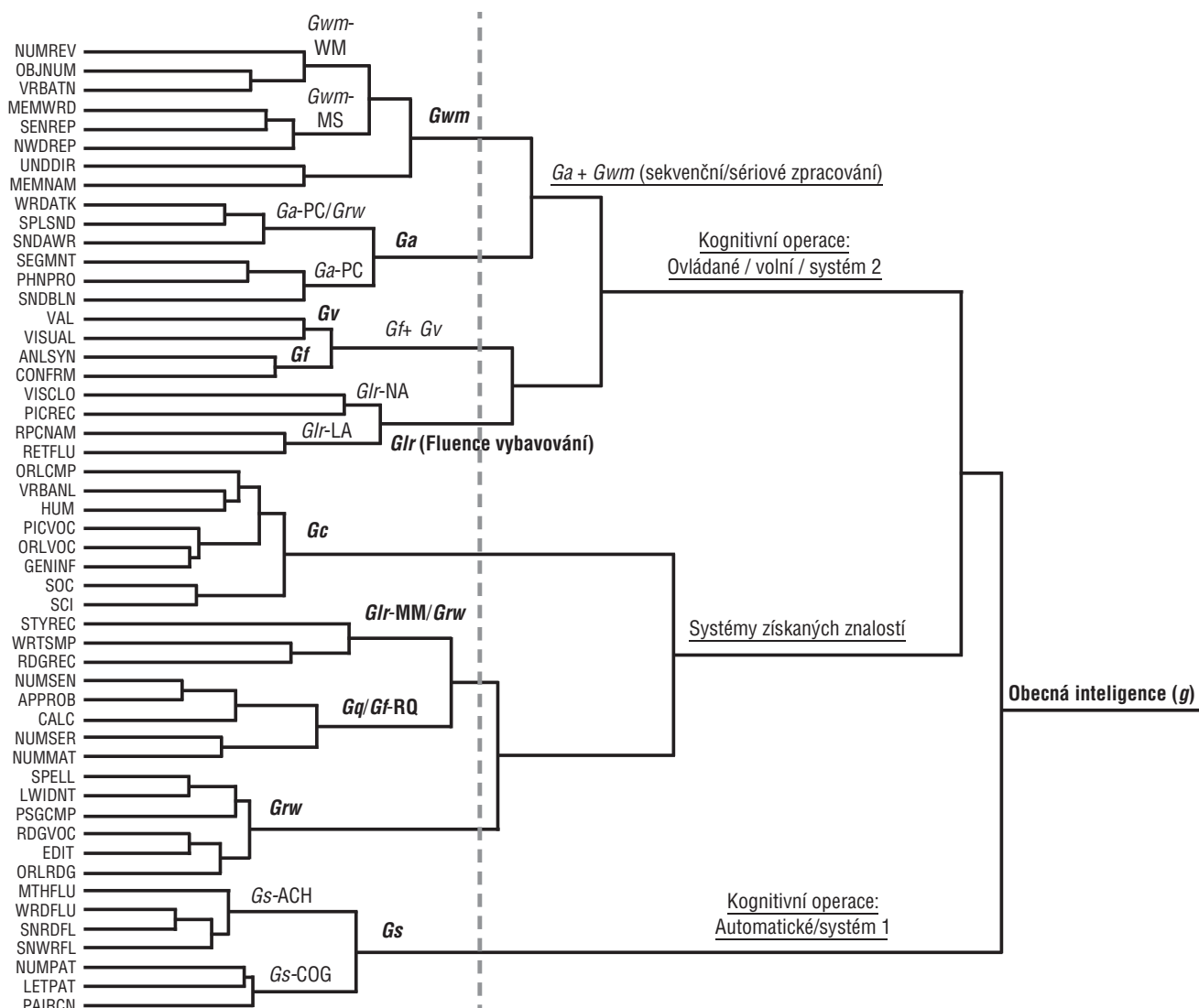
76 Kline (2011) někdy generování modelů (MG) nazývá také budováním modelů (*model building approach*).

77 Tento explorační přístup nebyl zcela čistý nebo slepý. Současná teorie CHC, jak je definovaná v kapitole 1 a Příloze A, byla použita jako výchozí teoretická optika při interpretaci výsledků generovaných těmito exploračními postupy. Důležité je, že rigidní model CHC nebyl jednoduše vnučen WJ IV ve výhradně konfirmační strategii CFA.

78 Zde přetištěn jako Obrázek 5-7.

Obrázek 5-7.

Wardova shluková analýza
pro věkové skupiny 9 až 13
let – výběrový soubor A pro
vývoj modelu ($n = 785$,
americká standardizace).



Pozn.: Tučný font = Široké nebo obecné schopnosti CHC. Podtržený text = Možné dimenze střední úrovně. Seskupení vlevo od vertikální čárkované čáry indikují, že jsou interpretovány jako validní schopnosti CHC. Seskupení vyššího řádu vlevo od čárkované linie jsou potenciálně zajímavé hypotetické dimenze schopností střední úrovně.

CA zahrnuje „třídění subjektů či proměnných na základě jejich podobnosti v jedné či více dimenzích a vytvoření skupin, které maximalizují vnitroskupinovou podobnost a minimalizují meziskupinovou podobnost“ (Henry, Tolan, & Gorman-Smith, 2005, s. 121). Literatura o CA zahrnuje široké spektrum různých třídících algoritmů a přístupů (Aldenderfer & Blashfield, 1984; Borgen & Barnett, 1987; Henry et al., 2005). Wardova hierarchická metoda minimálního rozptylu, která minimalizuje vnitrosklukovou sumu čtverců každého shluku při jejich spojení, je jednou z nejpoužívanějších empiricky prověřených metod CA pro odhalení již známé taxonomické struktury (Borgen & Barnett, 1987; Henry et al., 2005; Konold, Glutting, McDermott, Kush & Watking, 1999). V našem kontextu Wardova metoda CA začíná vzájemným propojením testů WJ IV, které jsou si navzájem nejbližší (tedy jsou nejvíce korelované), do skupin⁷⁹. Jakmile jsou tyto skupiny vytvořeny, propojují se dále s ostatními skupinami nebo jednotlivými testy, aby vytvořily skupiny větší. Jak je ilustrováno Obrázkem 5-7, tento proces pokračuje, dokud nejsou všechny testy spojeny v jediné všezahrnující skupině. Interpretace

⁷⁹ Abychom se vyhnuli záměně při užívání termínu „klastř“, který definuje specifickou kombinaci testů WJ IV, odkazujeme dále v této kapitole ke shlukům vyprodukovaným shlukovou analýzou také jako ke „skupinám“ či „uskupením“.

CA je poněkud subjektivní, ale jejího cíle je typicky dosaženo v bodě, kdy další spojování jednotlivých skupin činí interpretaci nadřazených skupin obtížnou, ne-li přímo nemožnou. Na Obrázku 5-7 tento bod prezentuje čárkovaná vertikála. Teoretické základy relevantní pro interpretaci výsledků CA vlevo od vertikály jsou objasněny v kapitole 1 a Příloze A.

Fáze 2A: Explorační analýza hlavních komponent (PCA)

Vztahy mezi 51 testovými proměnnými (skóry) byly rovněž zkoumány explorační analýzou hlavních komponent (PCA) s následnou rotací varimax (vytvářející kolmé nebo nekorelované komponenty). Přestože metoda společných faktorů nebo hlavních os následovaná šikmou rotací (pro korelované faktory) je typickou metodou explorační faktorové analýzy pro zkoumání teoretické struktury v sadě korelovaných testových proměnných a konstruktů, z několika důvodů byla použita PCA. Zaprvé, analýza hlavních komponent byla provedena v rámci explorační fáze vývoje modelu a její výsledky byly podpořeny dvěma dalšími exploračními metodami – shlukovou analýzou a multidimenzionálním škálováním – takže explorační PCA nebyla jedinou metodou užitou k exploraci dimenzionality baterie WJ IV. Zadruhé, kombinace exploračních výsledků byly užity primárně pro specifikaci nej přijatelnější struktury baterie WJ IV pro účely ověření metodami CFA. Zatřetí, korelace mezi latentními faktory jsou odhadnuty v navazující CFA. A konečně, užití PCA s rotací varimax bylo motivováno prakticky – analýza 51 testů, z nichž mnohé jsou vysoce korelované uvnitř, ale i napříč doménami CHC, představuje unikátní statistickou výzvu. Například doména Grw zahrnuje 13 testů čtení a psaní se středně silnými až silnými vzájemnými korelacemi. Ve třech bateriích WJ IV (COG, OL a ACH) je celkem 7 testů Gc a ne méně než 9 testů rychlosti nebo fluence (Gs). Z toho všeho vyplývá, že nebylo neobvyklé, že v případě nejméně jednoho (ale spíše více) testu jsou jeho skóry dokonale nebo téměř dokonale predikovatelné z lineární kombinace jiných vysoce korelovaných testů. Přítomnost takto vysoké multikolinearity⁸⁰, která neproblematizuje data nebo testy WJ IV, ale spíše je funkcí extenzivního pokrytí určitých domén v baterii WJ IV, může vést k problémům s konvergencí při EFA metodou hlavních os a následnou šikmou rotací. PCA s rotací varimax je výpočetně jednodušší a méně náchylná k tzv. nevlastním řešením (Brown, 2006; Duntelman, 1989). Použití této metody eliminovalo většinu problémů s multikolinearitou a umožnilo použít analýzu kompletní sady 51 testů. Oproti tomu šikmá řešení metodou hlavních os by vyžadovala iterativní selekci a vyřazování vždy jednoho testu v každém kroku, a to do doby, než by bylo odhadnuto stabilní šikmé řešení se společnými faktory. Tato ztráta informace o testech a faktorech plynoucí z odstranění WJ IV testů z analýzy pro nás nebyla akceptovatelná.

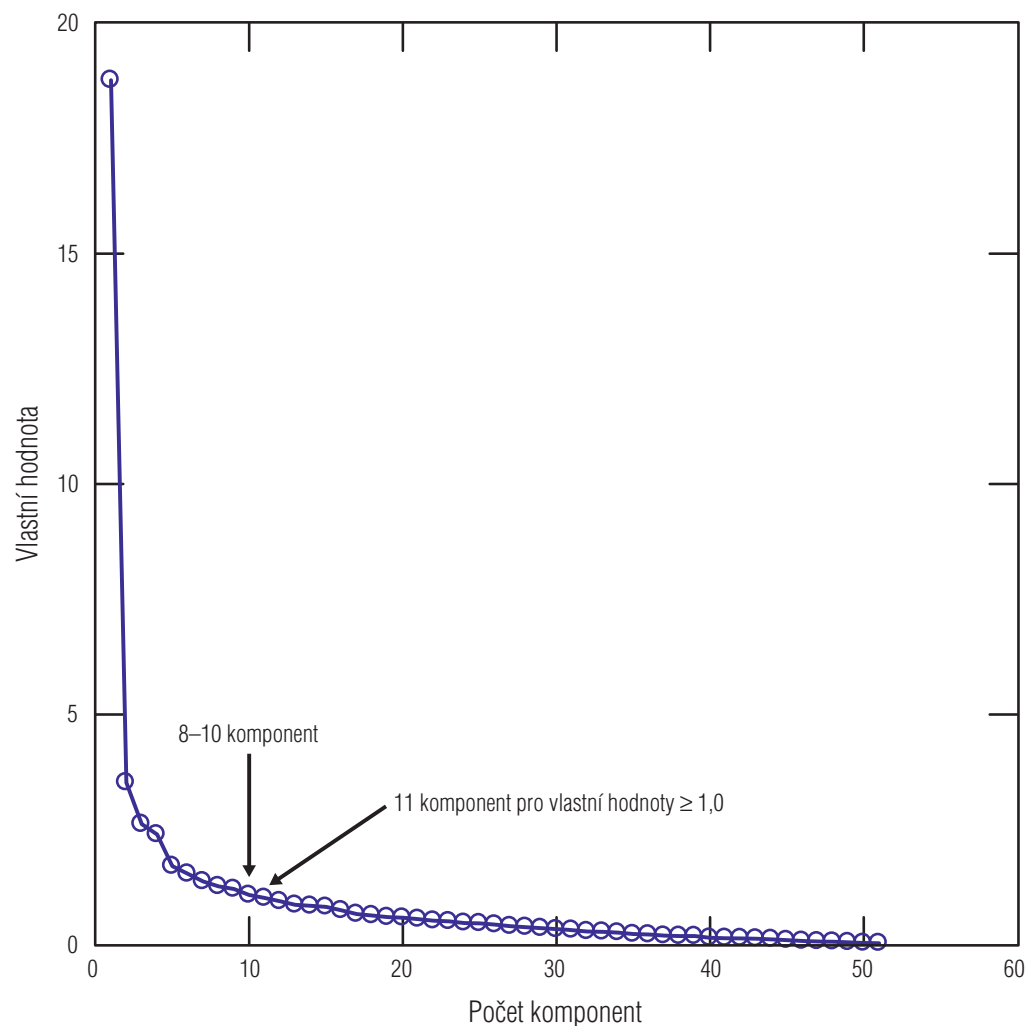
Jak poznamenal Brown (2006), striktní dodržování specifického kritéria extrakce faktorů nebo komponent není snadné; „*proces faktorové analýzy by měla silně řídit i věcná a praktická rozvaha*“ (s. 30). Vzhledem k explorační povaze PCA jsme pro určení počtu extrahovaných komponent v každém ze šesti souborů pro vývoj modelu využili jednoduché a flexibilní sady pravidel. Obrázek 5-11 v americkém technickém manuálu⁸¹ znázorňuje sutinový graf (scree plot; Cattell, 1966) vlastních hodnot odvozených z PCA sady 51 testů administrovaných souboru pro vývoj modelu ve věkové kohortě 9–13 let.

80 Přítomnost multikolinearity byla zjišťována z ukazatele *variable inflation factor* (VIF). Multikolinearita „se může objevit, přestože korelace mezi faktory nejsou excesivní“ (Keith, 2005, s. 200).

81 Zde přetištěný jako Obrázek 5-8.

Obrázek 5-8.

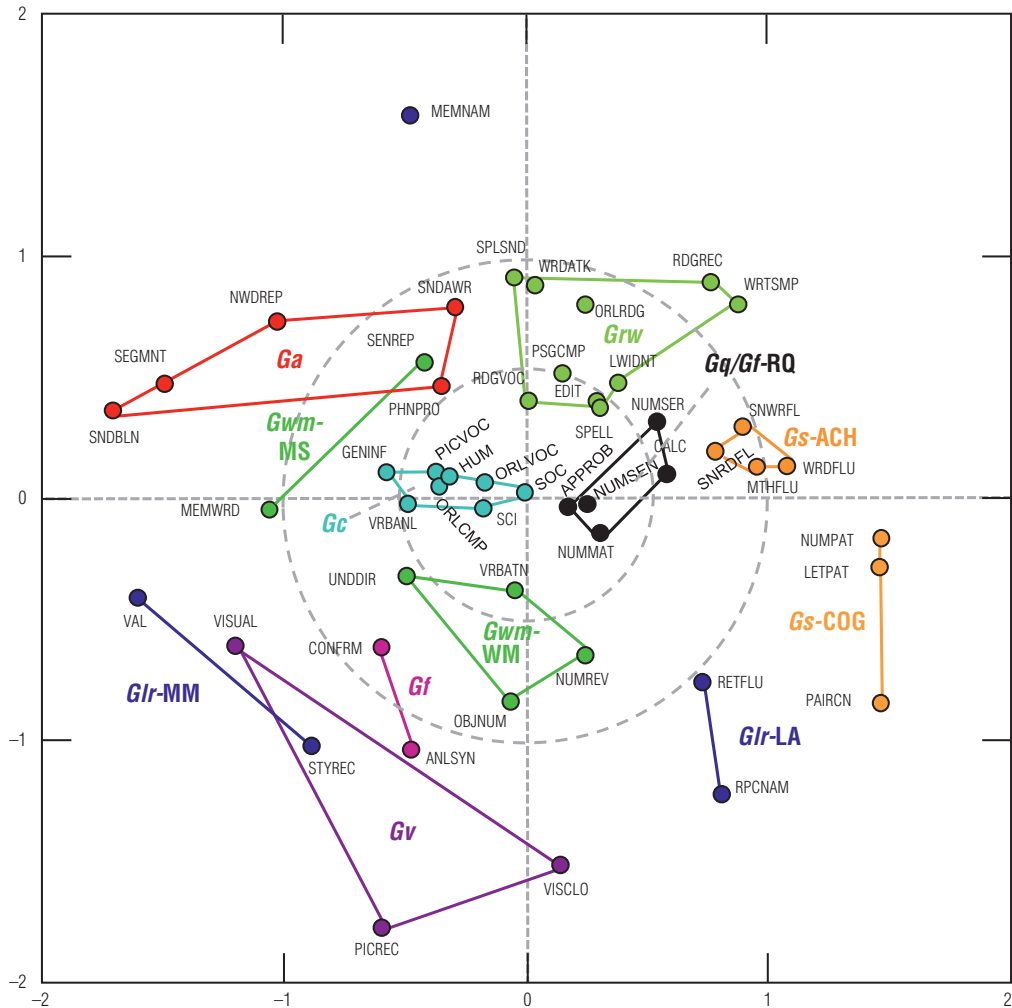
*Sutinový graf (vlastní hodnoty / latentní kořeny)
pro věky 9 až 13 let
– výběrový soubor A pro
vývoj modelu*

**Fáze 2A: Explorační multidimenzionální škálování**

Vztahy mezi 51 testy byly rovněž zkoumány metodou multidimenzionálního škálování (MDS). Popis MDS a jeho aplikace na standardizační data WJ IV je popsán dříve v této kapitole. Obrázek 5-9⁸² prezentuje výsledky na 51 testech administrovaných souboru pro vývoj modelu ve věkové kohortě 9–13 let.

82 V americkém originále jako Figure 5-12 na s. 162.

Obrázek 5-9.
MDS (Guttmanův radex) pro
WJ IV pro věky 9 až 13 let
let – výběrový soubor A pro
vývoj modelu (n = 785)



Pozn.: Ani v originále nejsou detailně vysvětleny zkratky. Zkratky pro širší schopnosti CHC by však měly být z předchozího textu srozumitelné.

Následující část budeme interpretovat vzhledem ke strukturní validitě. Jak jsme sdělili již v části této kapitoly věnované obsahové validitě, testy s vysokou prostorovou blízkostí na mapě MDS jsou interpretovány jako měřící podobné dimenze a testy, které jsou blíže středu grafu, jsou kognitivně komplexnější.

Fáze 2B: Specifikace počátečních modelů CFA

Zjištění z CA, PCA a MDS výběrového souboru pro vývoj modelu dětí ve věku 9-13 let byla v mnoha ohledech podobná, což poskytuje typ metodově invariantních důkazů o strukturní validitě mnoha širokých a vybraných úzkých schopností CHC. Nicméně výsledky CA a MDS nabízí do jisté míry odlišný vhled do struktury baterie WJ IV. Jak je znázorněno na obrázku 5-6, dalším krokem (Fáze 2B), byla specifikace počátečních modelů CFA založená na integraci výsledků CA, PCA a MDS. Jak je patrné z popisu fáze 2B, modely byly definovány i s přihlédnutím k podrobnému přehledu současných výzkumných poznatků v oblasti CHC a neurovědy (viz kapitola 1) a také k informacím o výzkumu struktury tří předešlých vydání testů Woodcock-Johnson (WJ, WJ-R a WJ III), které byly integrovány s výsledky exploračních analýz ve fázi 2A.

Fáze 3: Cross-validizace konfirmačních strukturních modelů

Zhodnocení fitu

Ke zhodnocení a srovnání různých modelů v různých věkových skupinách byla využita celá řada fit indexů CFA. Modely byly hodnoceny nejen s ohledem na celkový fit, ale i s přihlédnutím ke směru,

síle a signifikanci všech jejich parametrů a možnosti lokalizovat oblasti zasažené zkreslením (Brown, 2006).^{83 84}

Důkazy o vnitřní struktuře: výsledky a interpretace důkazů o strukturní validitě

Hodnocení modelů CFA a posouzení shody s daty nejsou pouze záležitostí statistiky:

„Cílem je v rámci třídy modelů nalézt takový, který se nejenom dobře shoduje s daty ze statistického úhlu pohledu tím, že bere v úvahu všechny možné zdroje chyb, ale má rovněž tu vlastnost, že každému z parametrů modelu lze dát věcně smysluplnou interpretaci.“ (Jöreskog, 1993, s. 307).

CFA modely WJ IV byly vyhodnocovány vzhledem k celkové shodě modelů s daty, a rovněž s ohledem na statistickou signifikanci a interpretabilitu všech odhadů parametrů modelu (Brown, 2006).

Srovnání shody modelů CFA s daty

Tabulka 5-14 v původním americkém technickém manuálu⁸⁵ prezentuje koeficienty shody modelu s daty finálních modelů CFA v podskupinách pro vývoj modelů a v podskupinách pro jejich validizaci v různých věkových skupinách.

Je k dispozici široká škála indexů shody (Arbuckle, 2012; Brown, 2006; Kline, 2011), kde každý z nich vysvětluje odlišné aspekty shody (např. míru podobnosti pozorované a modelem implikované variančně-kovarianční matice; úspornost modelu). Indexy shody reportované v tabulce 5-14 zahrnují (a) minimální diskrepanci (CMIN)⁸⁶ a stupně volnosti (df), adjustovaný index dobré shody (AGFI), (c) komparativní index shody (CFI), (d) Tuckerův-Lewisův nenormovaný index shody (TLI), (e) normovaný index shody adjustovaný vzhledem k úspornosti (PNFI) a (f) odmocninu z průměrného čtverce chyby aproximace (RMSEA)⁸⁷. Obecně vzato, čím blíže jsou AGFI, CFI, TLI a PNFI blíže k hodnotě 1, tím lepší je shoda modelu. V závislosti na specifickém indexu se požadované hodnoty typicky pohybují v rozmezí od 0,90 do 0,95 a výše. V případě indexu RMSEA naopak malé hodnoty blíží se k nule indikují dobrý fit modelu, hodnoty v rozmezí 0,05 až 0,08 typicky naznačují dobrou shodu modelu s daty. Tato obecně zažitá pravidla nicméně dále závisí na řadě různých charakteristik vzorku, dat a modelu (viz Brown, 2006; Fan, Thompson, & Wang, 1999; Hu & Bentler, 1999; Kline, 2011).

Vztah skóre WJ IV s jinými nástroji na měření kognitivních schopností⁸⁸

Nutné a dostačující podmínky pro dosažení konstruktové validity jsou naplněny tehdy, když strukturně validní nástroje vykazují očekávané konvergentní a divergentní vztahy s nástroji, které měří stejné nebo jiné konstrukty než zkoumaná metoda (Benson, 1998; Benson a Hagtvet, 1996; Cronbach a Meehl, 1955). Tato podkapitola pojednává o výsledcích studií, které zkoumaly vztahy mezi skóre WJ IV a řadou externích kritérií. Typy uvedených důkazů o externí validitě zahrnují souběžnou validitu WJ IV COG s jinými bateriemi k měření inteligence.

83 Lokalizované zkreslení modelu („*localized model strain*“) nastává ve chvíli, kdy některé vztahy mezi indikátory v datech nejsou reprodukovány adekvátně nebo některé z modelu plynoucí vztahy výrazně překračují vztahy pozorovatelné v datech. To se nejčastěji děje u komplexních modelů s vysokými korelacemi mezi indikátory (Brown, 2006).

84 Informace o použitých koeficientech shody jsou prezentovány v následujících sekcích této kapitoly („Důkazy o vnitřní struktuře: výsledky a interpretace důkazů o strukturní validitě“).

85 Je značně rozsáhlá a uvedena na s. 168-9. Zde ji nepřetiskujeme, pouze je namístě komentovat, že hodnoty v ní uvedené nenaznačují příliš přesvědčivou shodu modelů s daty, čemuž se ale na druhou stranu nelze příliš divit. Jak bude napsáno dále, plnohodnotná strukturní analýza dat WJ IV je značně náročný úkol, který na nás ještě čeká.

86 CMIN představuje hodnotu minima diskrepanční funkce pro každý model, v případě odhadu metodou maximum-likelihood (ML) se tedy jedná o hodnotu chi-kvadrátu (Arbuckle, 2012).

87 Indexy shody, které zde uvádíme, jsou obdobné, jaké obvykle reportují manuály jiných testů intelektu (viz Kaufman & Kaufman, 2004a; Roid, 2003b; Wechsler, 2008) a představují různé indexy z kategorie tzv. *přibližných* indexů shody (absolutních, inkrementálních nebo komparativních, adjustovaných vzhledem k úspornosti, nebo prediktivních; Kline 2011). AGFI vyjadřuje podíl variančně-kovarianční matice pozorovaných skóre vysvětlený modelem implikovanou variančně-kovarianční maticí, s korekcí na stupně volnosti. CFI s využitím nulového modelu indikuje, jak konkrétní model reprodukuje pozorovanou variančně-kovarianční matici v porovnání s modelem nulovým. TLI je podobný jako CFI, nicméně kontroluje stupně volnosti. PNFI rovněž kontroluje stupně volnosti, je nicméně založen na normovaném indexu shody (NFI); NFI a PNFI penalizují za komplexitu modelu ve prospěch úspornosti. RMSEA je odhadem míry, do níž souhlasí odhadovaná variančně-kovarianční matice a populační kovarianční matice, po korekcí na stupně volnosti. RMSEA je také někdy nazývána „index špatné shody“ (Brown, 2006; Kline, 2011; Mulaik et al, 1989).

88 Zbytek textu této kapitoly převzatého z amerického Technického manuálu přeložil Petr Pališek.

WJ IV staví na předchozím vydání, WJ III, jehož důkazy pro souběžnou validitu byly v nezávislé revizi Bradena a Niebelinga (2012) ohodnoceny téměř maximálním hodnocením (4 body z 5).

Studie souběžné validity byly provedeny na náhodně sebraných participantech tak, aby byl pokryt typický rozsah schopností americké populace. Tyto soubory neměly být, na rozdíl od normativního vzorku, reprezentativní pro tamní populaci. Není-li řečeno jinak, platí, že primární analytickou jednotkou studií je klastrový nebo složený skór. V tabulce 5-23 na s. 186 původního amerického manuálu jsou uvedeny velikosti vzorků (N), demografické informace každé ze studií, včetně informace o věku (s průměrem, směrodatnou odchylkou a rozsahem), a také procentuální zastoupení pohlaví, ras, etnicit a vzdělání matky a otce. V českém prostředí tyto studie zatím nebyly provedeny, proto je následující výsledky při aplikaci na tuzemský kontext nutné považovat za orientační.

Vztah originální verze WJ IV COG k jiným nástrojům měření inteligence byl prověřován v pěti studiích, které se zaměřovaly na následující nástroje: Wechslerova inteligenční škála pro děti, 4. vydání (WISC-IV) (Wechsler, 2003), Wechslerova inteligenční škála pro dospělé, 4. vydání (WAIS-IV) (Wechsler, 2008), Wechslerova škála inteligence pro předškolní a školní děti, 3. vydání (WPPSI-III) (Wechsler, 2002), Kaufmanova baterie pro vyšetření dětí, 2. vydání (KABC-II) (Kaufman a Kaufman, 2004a), Inteligenční škála Stanford-Binet, 5. vydání (SB-5) (Roid, 2003a) a Škála diferenciálních schopností, 2. vydání (DAS-II) (Elliott, 2007). Každá z těchto metod je individuálně administrovaným vyšetřením inteligence a kognitivních schopností.

Výsledky těchto srovnání jsou uvedeny v tabulkách 5-24, 5-25, 5-26, 5-27, 5-28 a 5-29 a jejich podrobný popis je k dispozici v manuálu pro originální verzi WJ IV (McGrew, LaForte, Schrank, 2014, s. 187-197). Souhrnně je možné o nich napsat, že dokládají převážně dobrou konvergentní validitu skóru WJ IV se skóry použitých metod. V tomto českém technickém manuálu je nepřetiskujeme, abychom čtenáře neuváděli v omyl, že jde o výsledky analýz dat české standardizace. Americký technický manuál je dostupný na webu, takže čtenáři, který by si chtěl výsledky vyhledat a porovnat s českými (i v budoucnu, až budou provedeny další validizační studie), nic nebrání.

Výkon ve WJ IV u osob z klinických vzorků

Jedním z důkazů o validitě mohou být korelace skóru WJ IV se skóry z testů, které měří podobné konstrukty. Dalším z důkazů pak může být vztah mezi skóry WJ IV a příslušností ke skupině (např. mezi osoby s poruchami učení nebo s poruchou intelektu). Autoři originální verze v manuálu uvádí (McGrew, LaForte, Schrank, 2014, s. 209), že vybrané testy administrovali lidem z těchto klinických skupin: nadání, poruchy intelektu/mentální retardace, poruchy učení, opoždění jazykového vývoje, ADHD, poranění hlavy a poruchy autistického spektra. Podrobné informace o tvorbě dat pro tato srovnání jsou k dispozici v originálním manuálu WJ IV (McGrew, LaForte, Schrank, 2014, s. 209-219). Je však třeba mít na paměti, že uvedené hodnoty a výstupy nelze bez dalšího výzkumu přenést do českého kontextu. Studie tohoto typu v našem prostředí ještě nebyly provedeny. Zmíněné informace proto mohou sloužit nanejvýš jako orientační údaj poukazující na potenciální vztah mezi českou verzí WJ IV a příslušností ke klinické podskupině. V žádném případě neodůvodňují využití WJ IV jako primárního zdroje pro diagnostiku ADHD, poruch autistického spektra a tak dále.

Údaje o srovnání skóru napříč zkoumanými klinickými podskupinami jsou uvedeny v tabulkách 5-40, 5-41 a 5-42 v originálním americkém technickém manuálu (McGrew, LaForte, Schrank, 2014).

Doklady strukturní validity založené na analýzách dat české standardizace⁸⁹

V následující kapitole jsou popsány důkazy o strukturní validitě české verze testu WJ IV COG. Ověření uvedených vlastností bylo provedeno na celém vzorku i na věkových segmentech, jež jsou blíže představeny níže. Struktura testu byla nejprve analyzována pomocí Pearsonova korelačního koeficientu mezi W-skóry všech dvojic testů a klastrových skóru. Dále byla provedena analýza hlav-

⁸⁹ Autory analýz jsou Edita Chvojková, Petr Pališek a Adam Tápal.

ních komponent, která je alternativou k faktorové analýze, již zde vzhledem ke komplexitě modelu a omezené velikosti vzorku nebylo možné využít. Závěrem je prezentována hierarchická shluková analýza spolu s multidimenzionálním škálováním, které mimo jiné umožňují vizualizaci struktury testu na základě shlukování testů s podobným vzorcem odpovědí.

Analýzy důkazů o validitě české adaptace testu WJ IV byly provedeny tak, aby byly podobné analýzám v původním americkém Technickém manuálu (McGrew, LaForte, Schrank, 2014). Při interpretaci výstupů z těchto analýz je nutné zohlednit, že jsou jejich možnosti omezeny tím, že je v české verzi nyní k dispozici jen kognitivní část testu WJ IV (COG), a není tedy možné prověřovat souvislosti s jazykovou (OL) a znalostní (ACH) částí, jak je tomu v původním technickém manuálu. Výpočty byly realizovány v programovacím jazyce R (R Core Team, 2019).

V této kapitole jsou v zájmu větší přehlednosti pro testy používány následující zkratky:

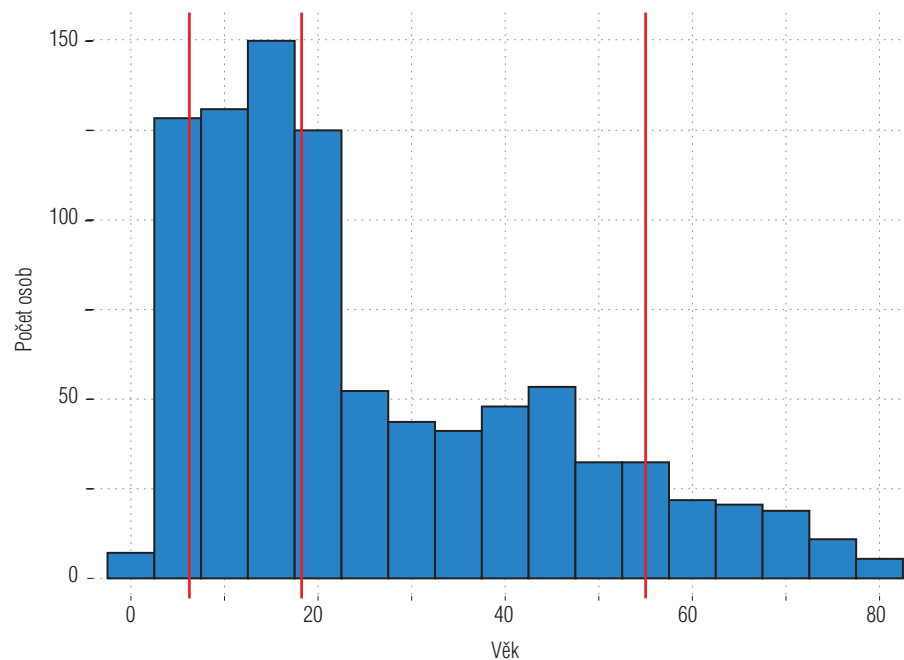
Tabulka 5-4.

Zkratky označující jednotlivé testy

Zkratka	Celý název
S	1. Slovník
ČŘ	2. Číselné řady
VP	3. Verbální pozornost
HP	4. Hledání písmen
FZ	5. Fonologické zpracování
RP	6. Reprodukce příběhů
V	7. Vizualizace
VZ	8. Všeobecné znalosti
FK	9. Formování konceptů
OŘ	10. Obrácené číselné řady
HČ	11. Hledání čísel
Pseu	12. Opakování pseudoslov
AVU	13. Audio-vizuální učení
RO	14. Rozpoznávání obrázků
AS	15. Analýza-syntéza
ŘNČ	16. Řazení názvů a čísel
VD	17. Vyhledávání dvojic
PnS	18. Paměť na slova

Výzkumný soubor (N = 918, průměr věku je 25,6 let, směrodatná odchylka věku je 18,9) byl za účelem zkoumání vývoje ve sledovaných charakteristikách v závislosti na stáří respondentů rozdělen do čtyř věkových segmentů (viz Tabulka 5-5 níže). Prahy byly voleny na základě východisek uvedených v technickém manuálu k původní verzi testu (McGrew, LaForte, Schrank, 2014; od s. 59 dále), jakož i s ohledem na to, aby velikost jednotlivých segmentů umožňovala provádění analýz (viz Obrázek 5-10, ve kterém vertikální přímky značí hranice mezi jednotlivými segmenty).

Obrázek 5-10.
Histogram věku participantů

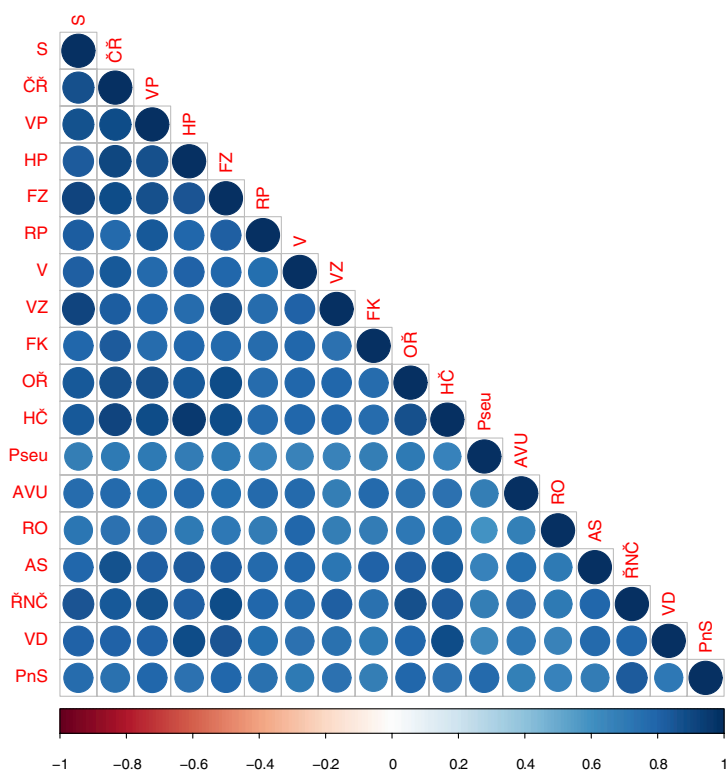


Tabulka 5-5.
Věkové segmenty používané při analýze

Věkový segment	Věkový interval	N
A	do 6,5 let (vč.)	103
B	od 6,5 let do 18 let (vč.)	333
C	Od 18 let do 55 let (vč.)	389
D	nad 55 let	93

V prvním kroku byla prověřována prostá asociace mezi W-skóry jednotlivých testů, k čemuž byl použit Pearsonův korelační koeficient. Výsledky jsou níže prezentovány prostřednictvím korelačních diagramů, které obecně vzato ukazují na vzájemnou souvislost výkonu v jednotlivých testech. Vysoká kolinearita všech testů dále předznamenává určité obtíže s odhadováním jednoznačné faktorové struktury. Tento aspekt je blíže prověřován v navazující, dosud nepublikované studii.

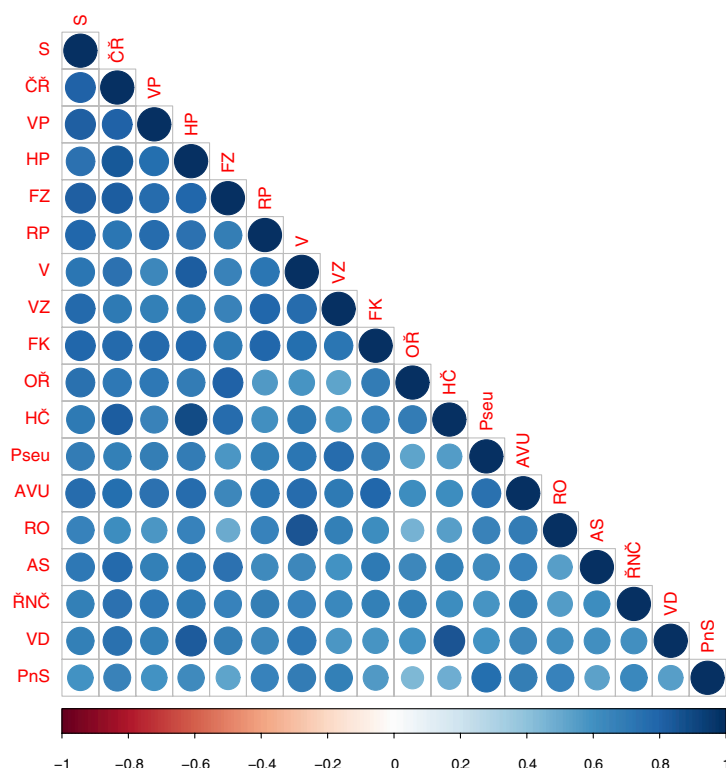
Obrázek 5-11.
Korelační diagram pro všechny testy na celém vzorku



Z vizuální inspekce korelačních diagramů je patrné, že se stoupajícím věkem slábnou souvislosti mezi jednotlivými testy (všechny korelace jsou však kladné), což je očekávatelné (vzhledem k narůstající variabilitě vývojových trajektorií). Na úrovni klastrových skórů (viz níže) je nejvýraznější postupné vydělování klastru Gs (Rychlost kognitivního zpracování), což interpretujeme nestejným poklesem obecného intelektu a rychlosti zpracování informací/psychomotorického tempa, popř. klastru Gv (Vizuální zpracování).

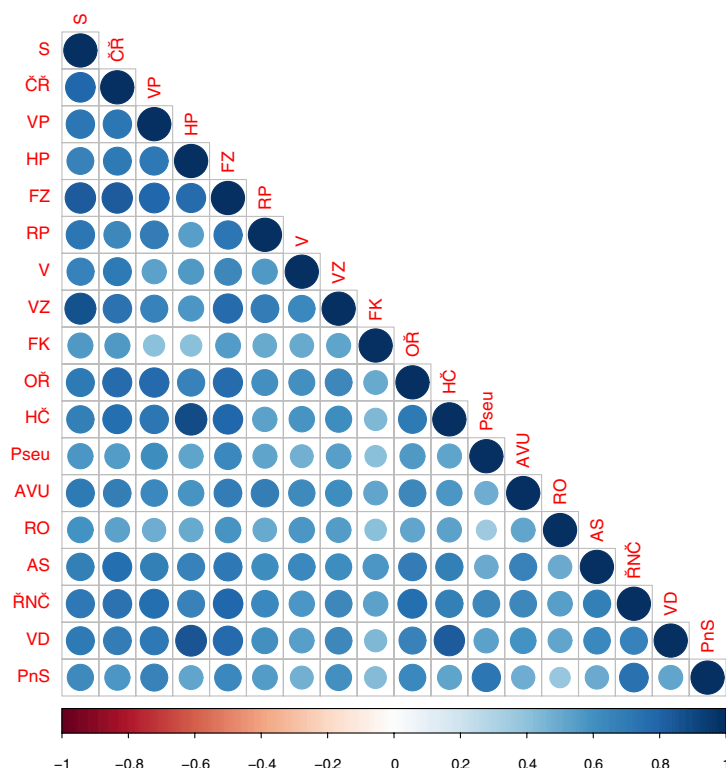
Obrázek 5-12.

Korelační diagram pro
W-skóry v jednotlivých
testech pro věkový segment
A



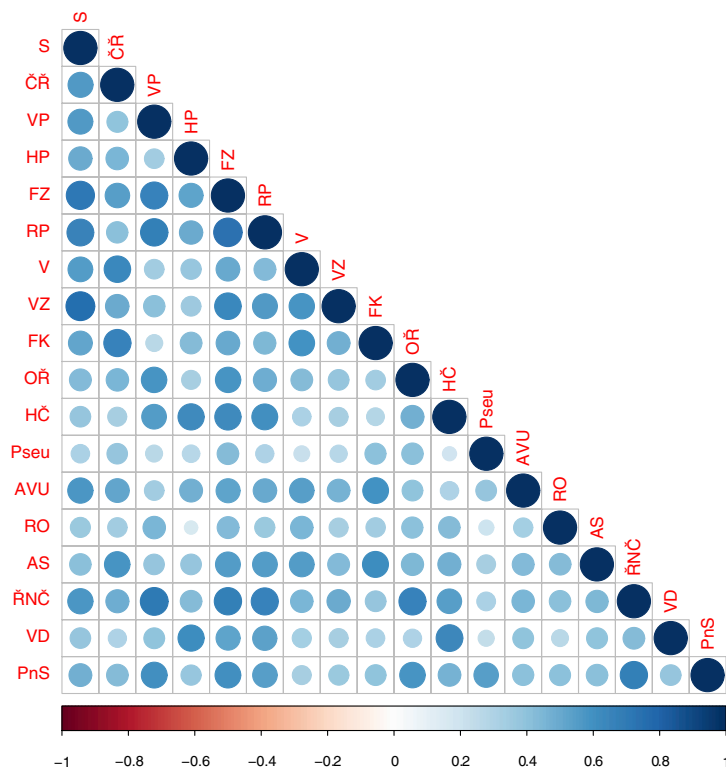
Obrázek 5-13.

Korelační diagram pro
W-skóry v jednotlivých
testech pro věkový segment
B



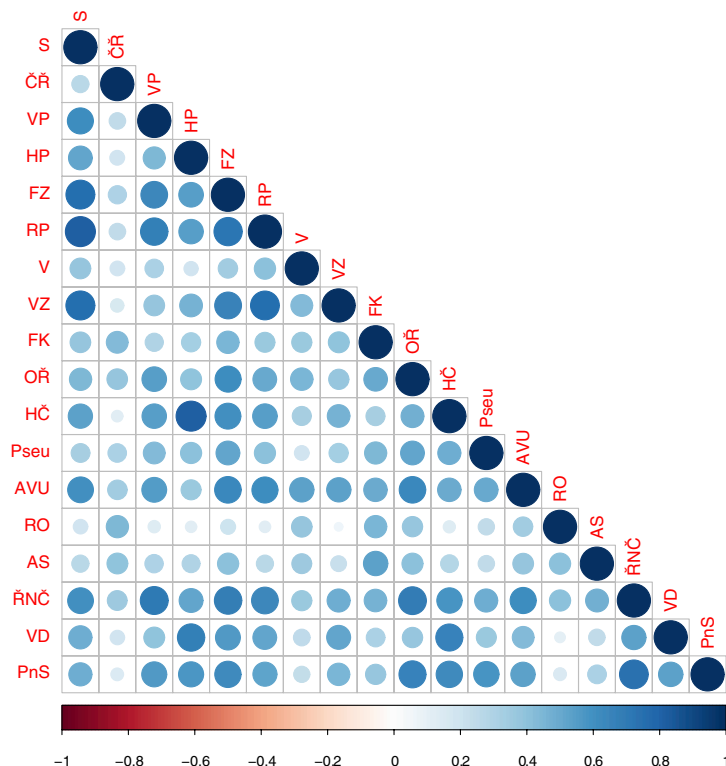
Obrázek 5-14.

Korelační diagram pro
W-skóry v jednotlivých
testech pro věkový segment
C



Obrázek 5-15.

Korelační diagram pro
W-skóry v jednotlivých
testech pro věkový segment
D

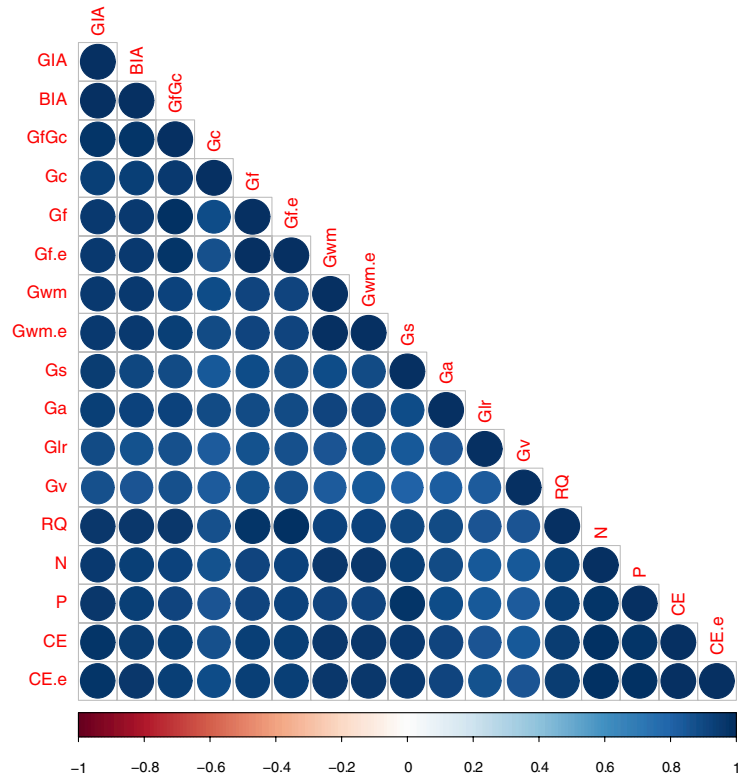


Obdobným způsobem byly prověřeny také asociace mezi klastrovými skóry (významy zkratk viz tab. 1-1). Výsledky jsou prezentovány níže a vybízí k obdobnému závěru o postupných změnách ve struktuře klastrových skóru – korelační vztahy s rostoucím věkem respondentů oslabují. Přesto zůstávají korelace jednotlivých klastrových skóru s obecnou inteligencí (klastř GIA) poměrně vysoké až do segmentu C (18-55 let včetně), teprve v segmentu D (nad 55 let) korelace některých klastrových skóru výrazněji oslabují (i když nadále jsou všechny kladné). Dá se říci, že i v nejstarší skupině respondentů se skórem GIA, které reprezentuje obecnou inteligenci, vysoce korelují klastrové skóry Porozumění – Znalosti (Gc),

Krátkodobá pracovní paměť (Gwm), Rychlost kognitivního zpracování (Gs), Auditivní zpracování (Ga), Dlouhodobá paměť (Glr), Rychlost vnímání (P) a Kognitivní efektivita (CE). Poněkud slabší vztah skóru GIA nacházíme u této věkové podskupiny se skórem Početních dovedností (N) a Kvantitativního uvažování (RQ), vůbec nejslabší pak s Fluidní inteligencí (Gf) a Vizuálním zpracováním (Gv). Při uvažování o těchto vztazích ale nesmíme ztrácet ze zřetele fakt, že nejde o longitudinální, ale průřezová data, oslabení těchto vztahů tedy nelze automaticky považovat za vývojový trend u všech respondentů.

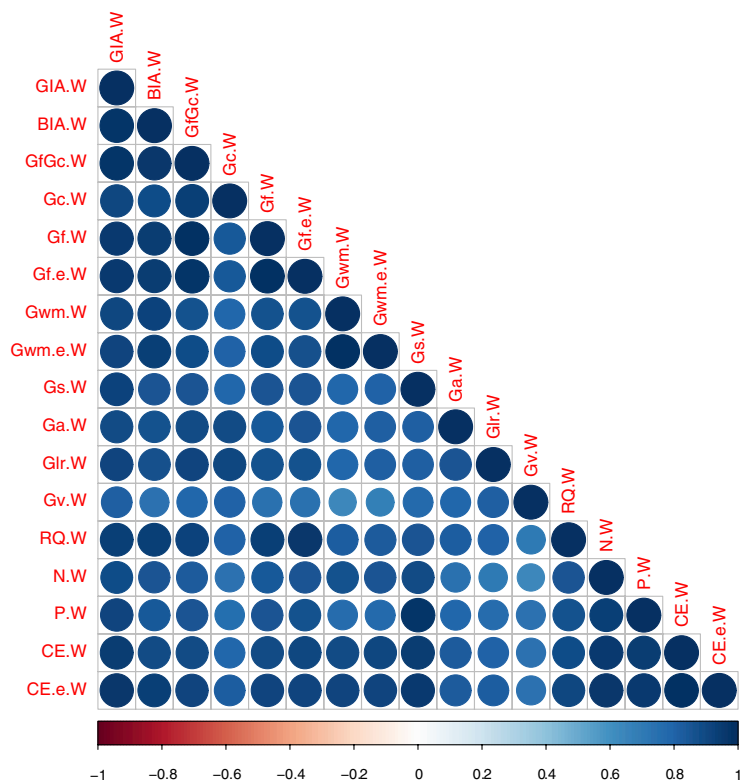
Obrázek 5-16.

Korelační diagram pro
klastrové skóry na celém
vzorku



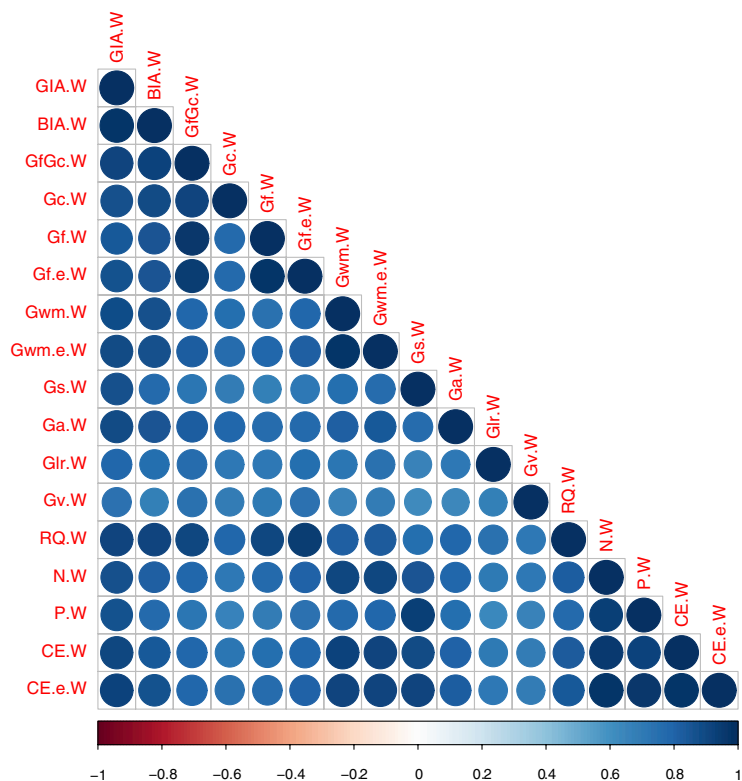
Obrázek 5-17.

Korelační diagram pro
klastrové skóry pro věkový
segment A



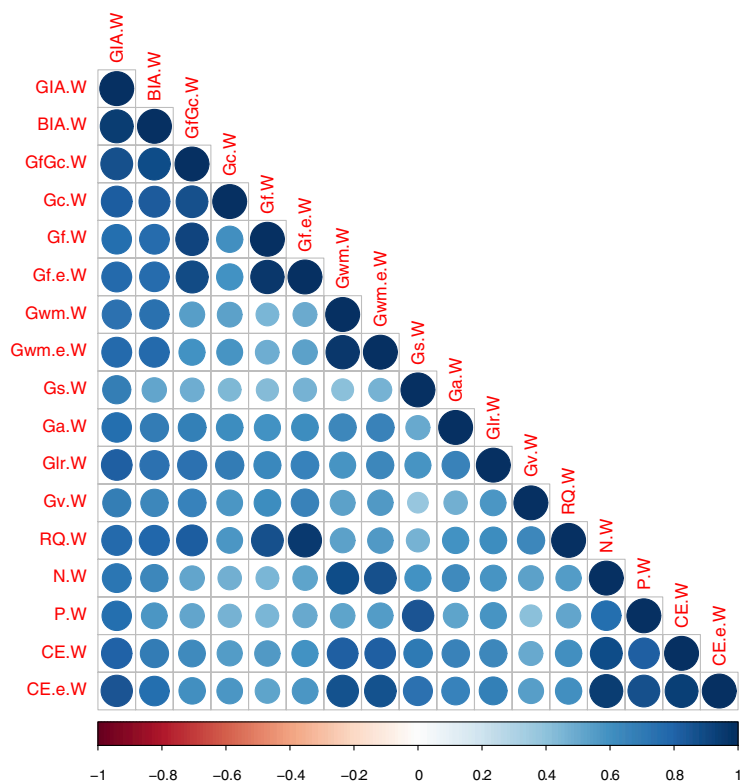
Obrázek 5-18.

Korelační diagram pro
klastrové skóry pro věkový
segment B



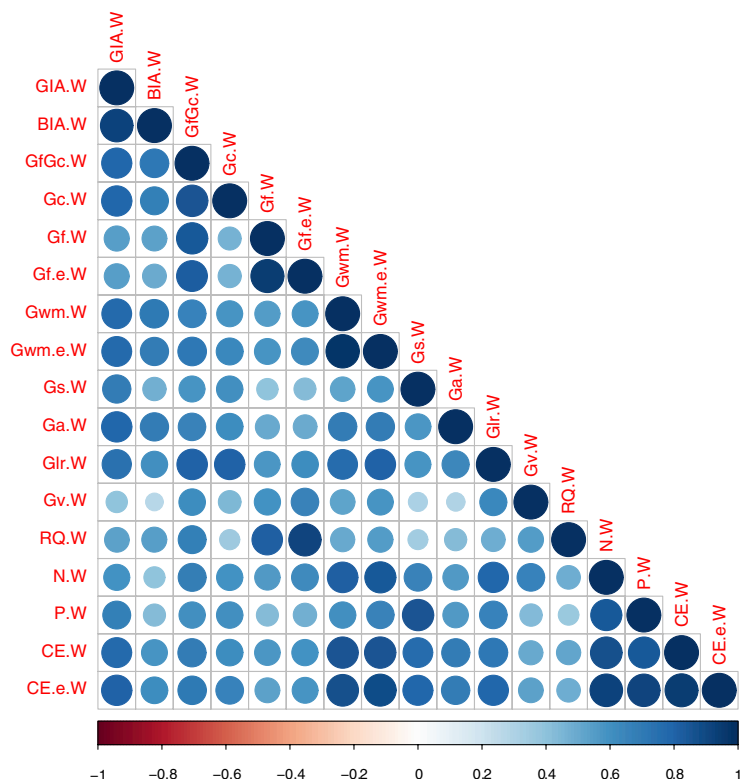
Obrázek 5-19.

Korelační diagram pro
klastrové skóry pro věkový
segment C



Obrázek 5-20.

Korelační diagram pro
klastrové skóry pro věkový
segment D



V ideálním případě by po inspekci korelačních matic následovalo provedení faktorové analýzy a zhodnocení souladu teorií implikovaného modelu s vytvořenými daty. Takový postup by nicméně vyžadoval násobně rozsáhlejší výzkumný soubor, než který byl při ověřování strukturní validity testu získán. Zvolili jsme proto alternativní postup spočívající v provedení analýzy hlavních komponent, která nemodeluje latentní proměnné, čímž snižuje nároky na velikost výzkumného souboru.

Na základě paralelní analýzy jsme postupně volili jedno- až tříkomponentové řešení PCA (bez Kaiserovy normalizace) s cílem po vzoru původního technického manuálu nalézt teoreticky uspokojivé vysvětlení vztahů mezi jednotlivými testy. Scree ploty pro jednotlivé věkové segmenty poukázaly na vhodnost dvoukomponentového řešení (zejména v segmentu D). Jednokomponentové řešení uspokojivě ukazovalo na existenci g-faktoru. Dvoukomponentové řešení pak zachovává primární komponentu (g-faktor) a zároveň koresponduje s výsledky shlukové analýzy (viz níže), protože od sebe odděluje testy, které ve shlukové analýze na nejvyšší úrovni spadají od dvou odlišných shluků. Trojkomponentové řešení se nejevilo jako vhodné, protože odhadnutá struktura (vztahy mezi komponentami a testy) neodpovídala teoretickým předpokladům. Problematickým se zde ukazoval být test Vyhledávání dvojic (VD), který bude blíže analyzován v navazujících studiích.

Tabulka 5-6.
Náboje jednotlivých
testů pro PCA s 1-3
komponentami

Zkratka testu	Komponenta	Komponenty		Komponenty		
S	0,93	0,78	0,51	0,70	0,46	0,41
ČŘ	0,95	0,84	0,44	0,70	0,57	0,32
VP	0,93	0,77	0,52	0,71	0,45	0,42
HP	0,93	0,85	0,40	0,73	0,52	0,29
FZ	0,94	0,79	0,52	0,76	0,41	0,42
RP	0,88	0,69	0,55	0,55	0,53	0,44
V	0,89	0,77	0,44	0,49	0,72	0,30
VZ	0,88	0,72	0,51	0,66	0,42	0,42
FK	0,86	0,72	0,48	0,43	0,72	0,34
OR	0,92	0,76	0,52	0,70	0,44	0,42
HČ	0,93	0,88	0,37	0,80	0,47	0,27
Pseu	0,78	0,35	0,86	0,28	0,40	0,79
AVU	0,84	0,66	0,53	0,38	0,71	0,40
RO	0,80	0,67	0,44	0,38	0,71	0,31
AS	0,89	0,80	0,41	0,59	0,63	0,28
ŘNČ	0,92	0,70	0,59	0,69	0,37	0,51
VD	0,88	0,80	0,38	0,77	0,38	0,29
PnS	0,85	0,47	0,81	0,50	0,29	0,75

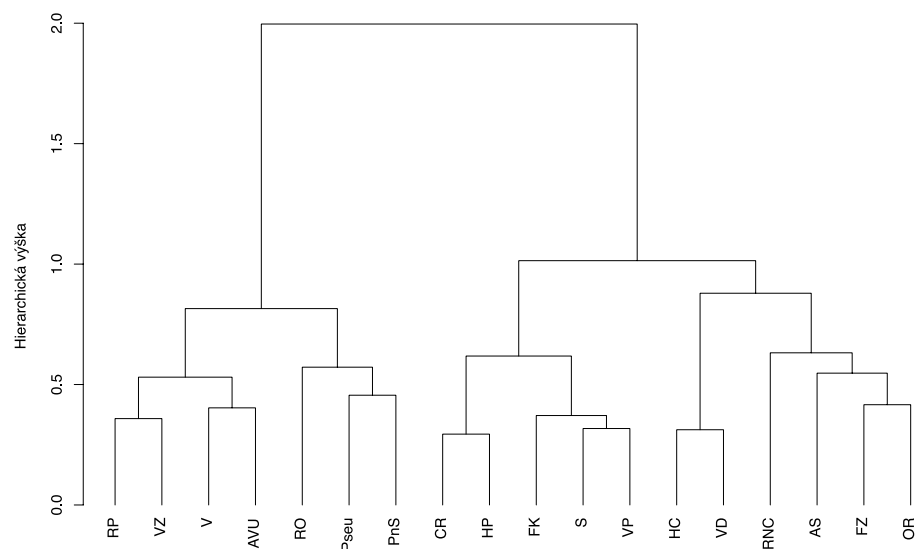
Hierarchická shluková analýza

Níže vykreslené dendrogramy popisují výsledky hierarchické shlukové analýzy pro každou z podskupin (odhadováno pomocí Wardovy metody na euklidovských distancích). Jde o další možnost, jak zkoumat strukturu vztahů mezi skóry jednotlivých testů. Prohlídka jednotlivých dendrogramů nepřináší jednoznačné potvrzení struktury korespondující s teorií C-H-C. Např. sedm testů tvořících klastrový skór GIA (tzn. zkratky S, ČŘ, VP, HP, FZ, RP a V) jsou ve všech čtyřech dendrogramech rozptýleny víceméně rovnoměrně. Na druhou stranu od testů tvořících odhad faktoru g bychom něco podobného očekávat mohli. Proto se podíváme na testy tvořící některé specifické klastry.

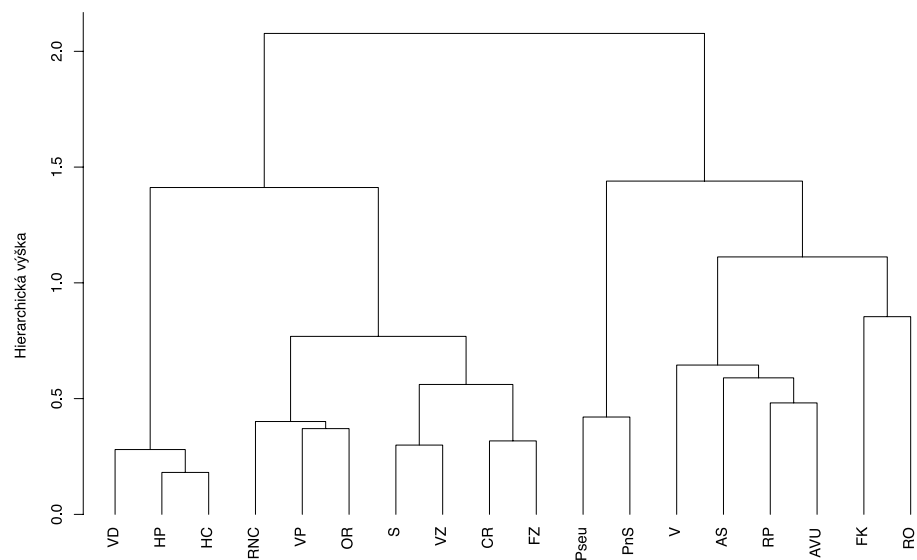
U obecnějších klastrů reprezentovaných skóry BIA nebo *Gf-Gc* je rozmístění jednotlivých testů v dendrogramech podobné jako v případě testů tvořících klastrový skór GIA. Teprve když se zabýváme klastry tvořenými dvojicemi testů (viz tab. 1-1), nacházíme podporu pro strukturu teorie C-H-C. Testy klastru Porozumění – Znalosti (*Gc*), kterými jsou Slovník (S) a Všeobecné znalosti (VZ), jsou kromě nejmladšího věkového segmentu A spojeny na nejtěsnější (segmenty B a C) nebo velmi těsné úrovni (segment D), testy klastru Fluidní inteligence (*Gf*), kterými jsou Číselné řady (ČŘ) a Formování konceptů (FK), jsou na nejtěsnější úrovni spojené v dendrogramu pouze v segmentu C, ale v dalších dvou segmentech (A a D) jsou spojeny velmi těsně, a pouze v segmentu B jen velmi volně. Testy klastru Krátkodobá pracovní paměť (*Gwm*), kterými jsou Verbální pozornost (VP) a Obrácené číselné řady (OR), jsou také součástí společných shluků v dendrogramu, i když ne vždy na úplně nejtěsnější úrovni, testy klastru Rychlost kognitivního zpracování (Hledání písmen – HP a Vyhledávání dvojic – VD) jsou ve všech čtyřech dendrogramech spolehlivě propojeny, naproti tomu testy Fonologické zpracování (FZ) a Opakování pseudoslov (Pseu), které tvoří klastr Auditivní zpracování (*Ga*), se přesvědčivě nesdružují ani v jednom z dendrogramů. Testy Reprodukce příběhů (RP) a Audio-vizuální učení (AVU), které tvoří klastr Dlouhodobé paměti (*Glr*), se dobře sdružují v segmentech A a B, hůře v segmentu D a téměř vůbec v segmentu C. Testy Vizualizace (V) a Rozpoznávání obrázků (RO) z klastru Vizuálního zpracování (*Gv*) sice nejsou spojeny v jednotlivých dendrogramech na úplně nejtěsnější úrovni, ale jsou spojeny dosti těsně.

Mohli bychom podobně pokračovat s dalšími klastrovými skóry a někde bychom našli přesvědčivější a někde méně přesvědčivou podporu pro strukturu testu reprezentovanou klastrovými skóry zastupujícími jednotlivé faktory C-H-C.

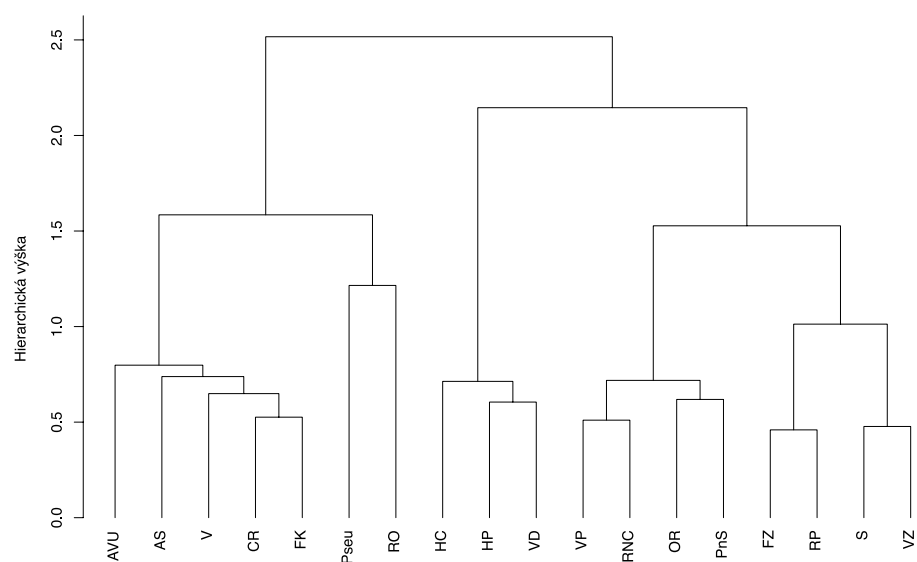
Obrázek 5-21.
*Hierarchická shluková
analýza pro věkový segment
A*



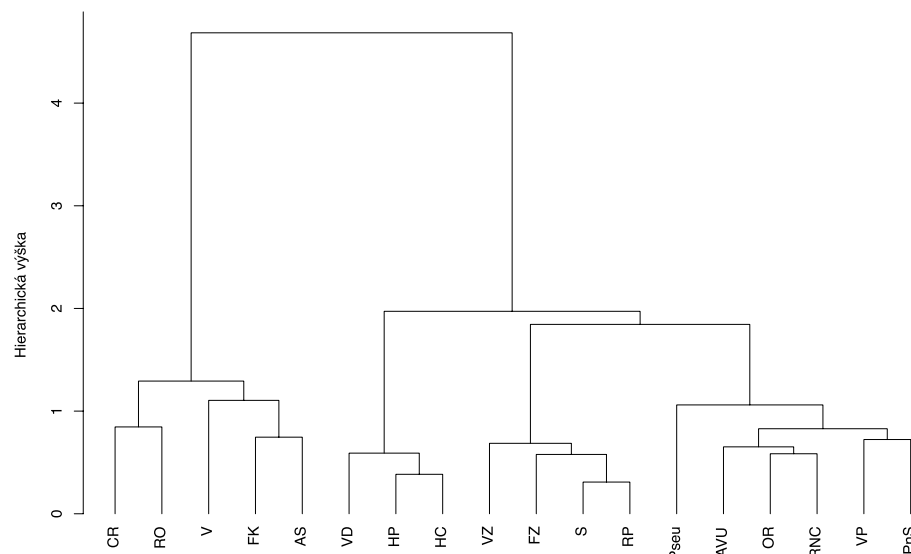
Obrázek 5-22.
*Hierarchická shluková
analýza pro věkový segment
B*



Obrázek 5-23.
*Hierarchická shluková
analýza pro věkový segment
C*



Obrázek 5-24.
Hierarchická shluková
analýza pro věkový segment
D

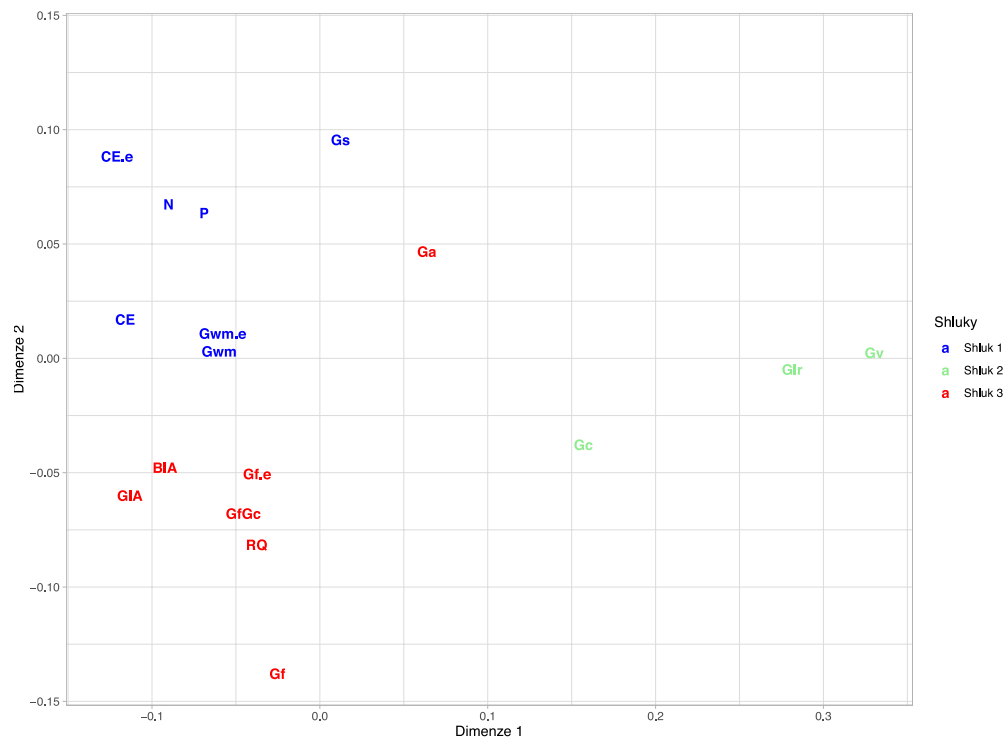


Multidimenzionální škálování

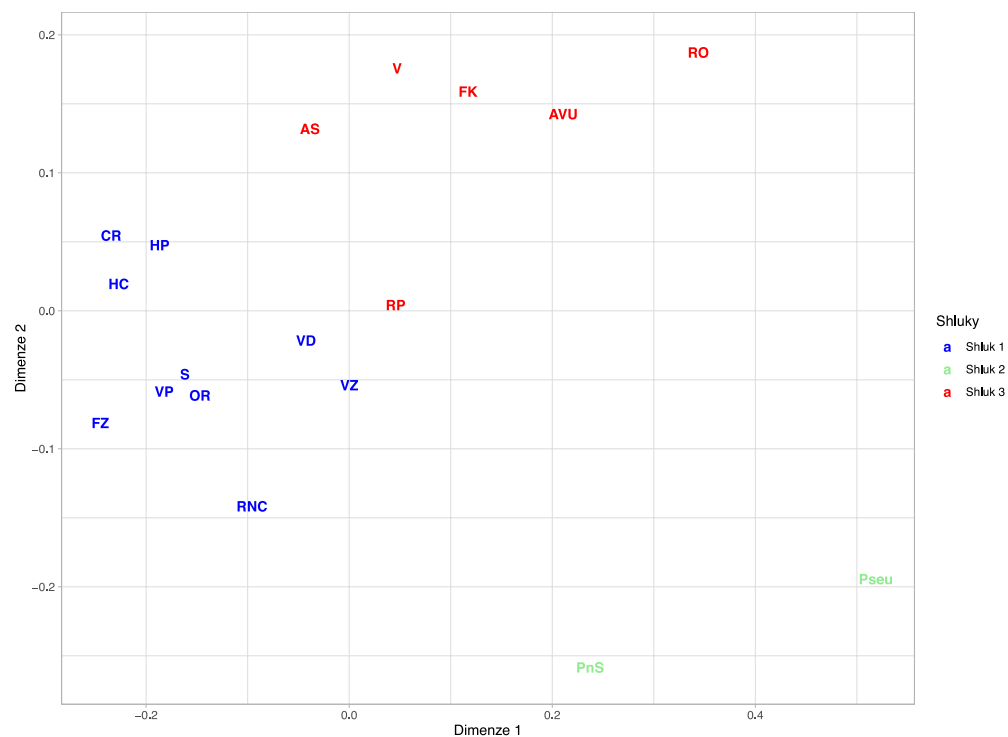
Provedli jsme rovněž multidimenzionální škálování (opět metodou euklidovských vzdáleností na korelační matici). Následují grafy pro škálování klastrů a škálování testů pro každou podskupinu. Shluky v MDS jsou tvořeny pomocí metody k-means ($k = 3$) přímo v analýze MDS a nejsou totožné se shluky v předchozí části.

Interpretace výsledných grafů je možné provádět na základě blízkosti nebo vzdálenosti umístění jednotlivých zkratků názvů testů a klastrových skóre. Obě osy dvourozměrného zobrazení je v případě těchto analýz obtížné interpretovat, protože by bylo vhodnější analyzovat skóre testů a klastrů WJ IV spolu s ϵ skóre dalších metod s odlišným obsahem. Vysoká kolinearita zmíněná už výše u korelačních diagramů zjevně neumožňuje dostatečně odlišit testy a klastry, které jsou vysoce podobné z hlediska odpovědí, od těch, které jsou pouze mírně podobné. Budeme tedy muset počkat na výzkumy obsahující další data, která podobným strukturním analýzám poskytnou širší kontext.

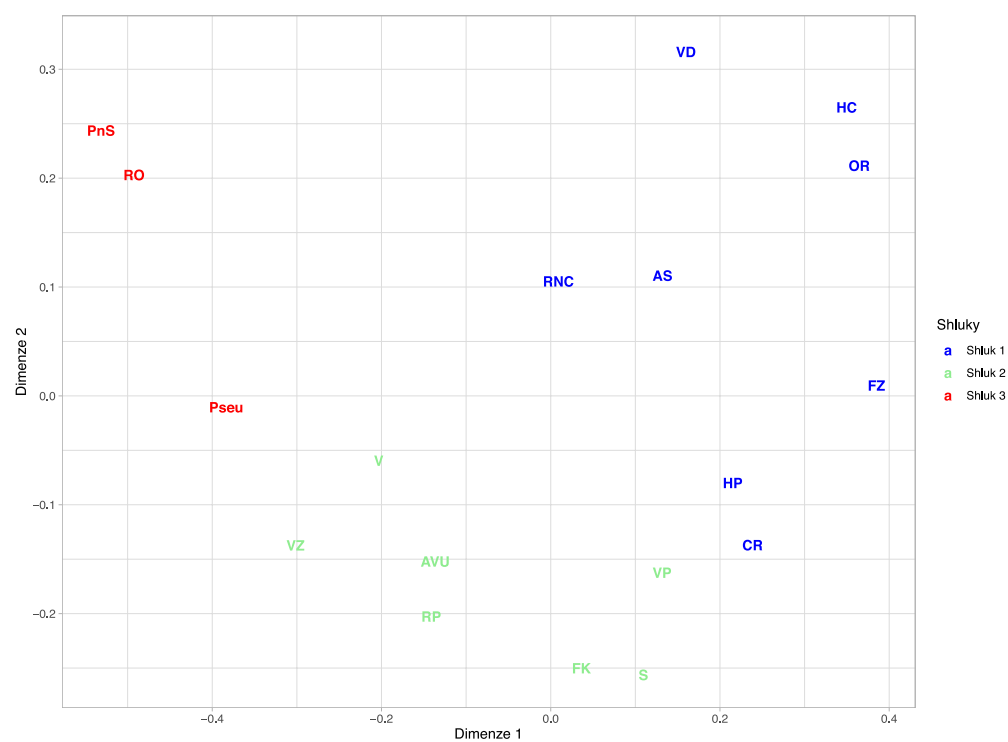
Obrázek 5-25.
MDS pro W-skóre klastrů
pro celý soubor



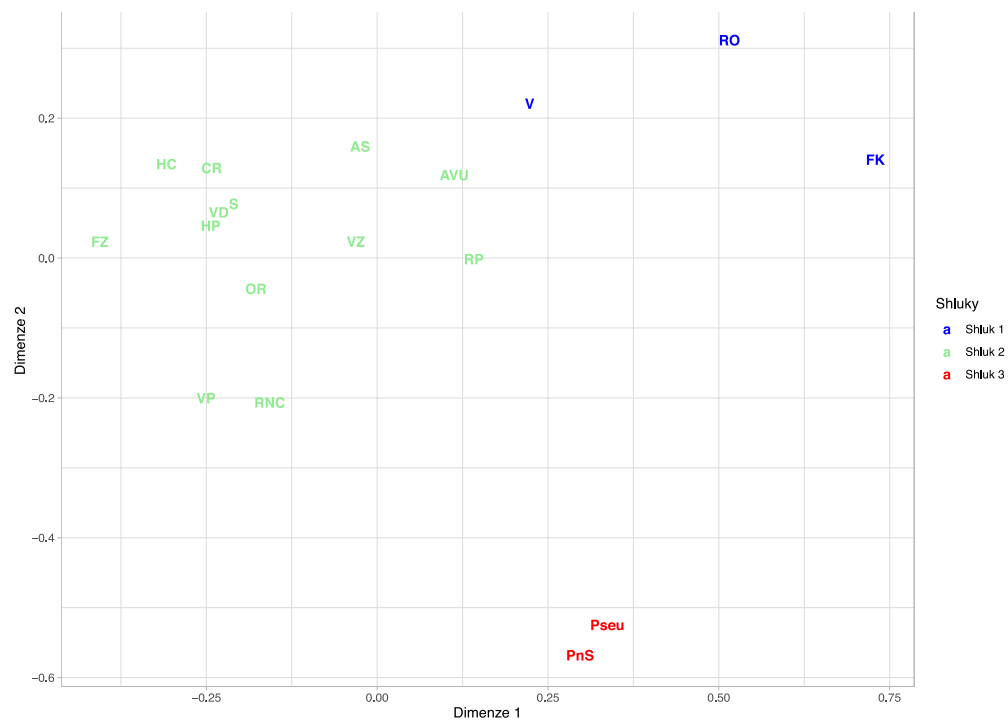
Obrázek 5-26.
MDS pro W-skóry testů pro
celý soubor



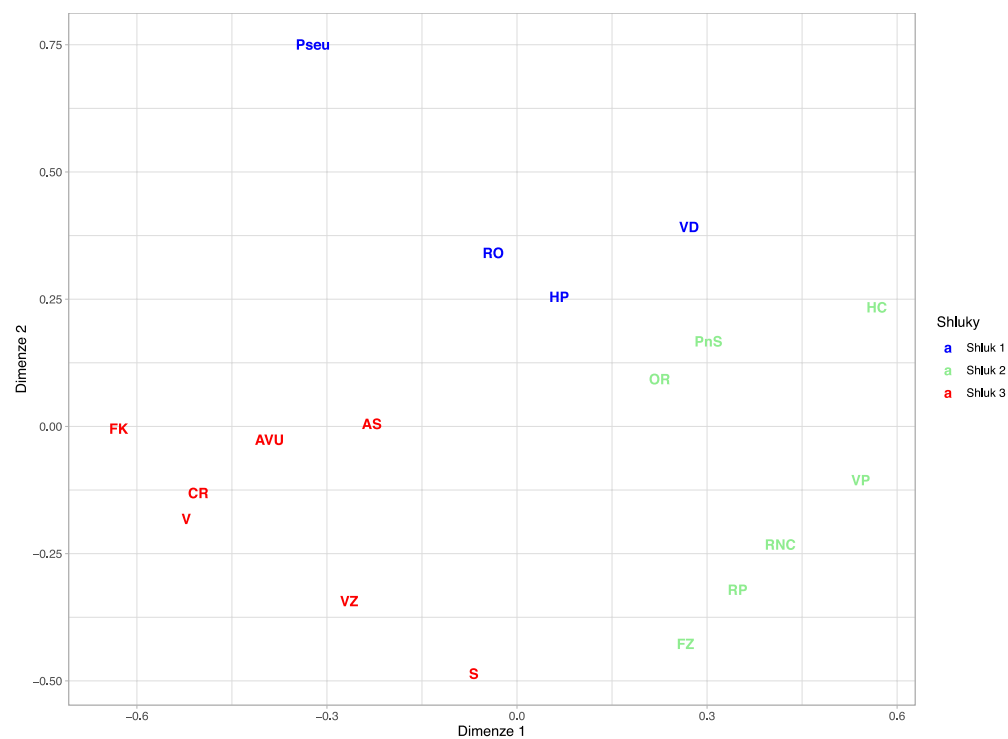
Obrázek 5-27.
MDS pro W-skóry klastrů
pro věkový segment A



Obrázek 5-28.
MDS pro W-skóry klastrů
pro věkový segment B

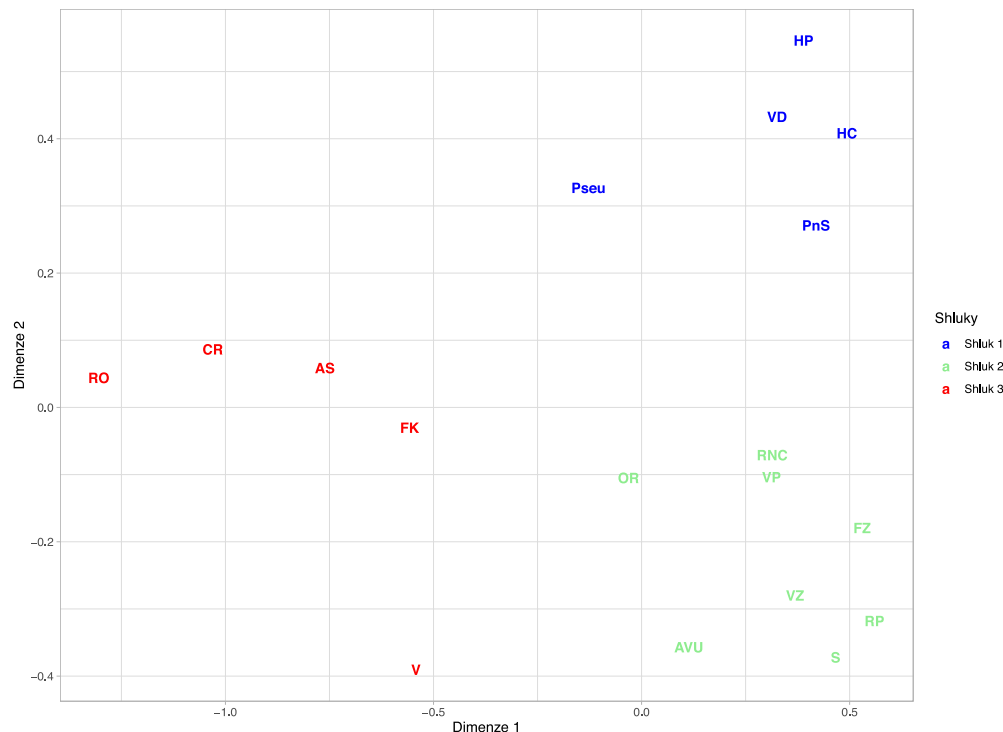


Obrázek 5-29.
MDS pro W-skóry klastrů
pro věkový segment C



Obrázek 5-30.

MDS pro W-skóry klastrů
pro věkový segment D



Validizační studie pro vybrané klinické skupiny⁹⁰

Test WJ IV COG byl administrován vybraným klinickým skupinám k rozšíření údajů o validitě také na vzorku lidí s konkrétními klinickými obtížemi (úzkost, deprese a schizofrenie). Sběr dat proběhl v různých psychiatrických nemocnicích a léčebnách v České republice, ale také mimo klinická zařízení. Po konzultaci s klinickým psychologem/psychiatrem byli zahrnuti respondenti s kýženou klinickou diagnózou klasifikovanou podle MKN-10: Mezinárodní klasifikace nemocí (WHO, 2016), kteří se přihlásili k účasti na výzkumu dobrovolně. Každá klinická skupina byla testována v rámci samostatné psychologické diplomové práce a měla mírně odlišný design (následující text bude strukturovaný vždy ve stejném pořadí: studie zabývající se úzkostnými poruchami, studie zabývající se depresivními poruchami a následně studie zabývající se schizofrenními poruchami).

Výběrové soubory:

Studii zaměřující se na *spektrum úzkostných poruch (F40-F49)* realizoval Mgr. Adam Klocek (Klocek, 2019). Souboru 26 respondentů z klinické skupiny bylo administrováno všech 18 testů. Klinické diagnózy, které se ve vzorku objevily, byly: smíšená úzkostně-depresivní porucha (46,2 %), obecná úzkostná porucha (11,5 %), panická porucha (11,5 %). Celkově 77 % lidí s komorbiditou nejen s dalšími neurotickými nebo somatoformními poruchami. Smíšená úzkostně-depresivní porucha byla včetně komorbidit zastoupena v 57,6 %. Nejčastějšími reportovanými způsoby medikace byly Miraklide, Mirtazapin, Rivotril, Zolof a Neurol.

Studii zaměřující se na *spektrum depresivních poruch (F30-F39)* realizovala Mgr. Martina Vítková (Vítková, 2017). Souboru 30 lidí byly administrovány pouze vybrané testy: Číselné řady, Formování konceptů, Analýza – syntéza, Reprodukce příběhů, Audio-vizuální učení, Vizualizace, Rozpoznávání obrázků, Hledání písmen, Hledání čísel, Vyhledávání dvojic. Klinické diagnózy, které se ve vzorku objevily: spektrum depresivních poruch (F32; N = 23), smíšená úzkostně-depresivní porucha (F41.2; N = 4), bipolárně-afektivní porucha s depresivní fází (F31.3/F31.4/F31.5; N = 2), smíšené a jiné poruchy osobnosti (F61; N = 4). Nejčastějším způsobem medikace byla antidepresiva, více než 50 % lidí užívalo také anxiolytika a několik také antipsychotika.

⁹⁰ Autory této kapitoly jsou Mgr. Adam Klocek, Mgr. Martina Schimová, Mgr. Martina Vítková, Mgr. Hynek Cígler, Ph.D.

Studii zaměřující se na *spektrum schizotypálních poruch a schizofrenie (F20.0-F20.9)* realizovala Mgr. Martina Schimová (Schimová, 2017). Vzorku 30 osob byly administrovány stejné vybrané testy jako v případě studie Vítkové (2017). Klinické diagnózy, které se ve vzorku objevily: paranoidní schizofrenie, hebefrenní schizofrenie, katatonní schizofrenie, nediferencovaná schizofrenie, postschizofrenní deprese, reziduální schizofrenie, schizophrenia simplex, jiná schizofrenie a schizofrenie NS. Medikace nebyla bohužel reportována.

Dohromady o všech třech diplomových pracích platí: mateřský jazyk čeština; národnost česká; celkové rozložení mužů (48 %) a žen (52 %) bylo napříč studiemi přibližně vyrovnané. Průměrný věk všech respondentů napříč studiemi byl 40,21 let. Data lidí z klinických skupin byla párována s daty lidí z kontrolní skupiny podle věku, pohlaví a vzdělání.

Dále uváděná doporučení jsou založena na studiu literatury, závěrečném rozhovoru s člověkem z klinické skupiny (případně v některých případech také na reflexi odborného pracovníka, který měl pacienta v péči) a na vlastním pozorování během administrace. Přesto by doporučení měla být zvažována s respektem k jejich orientační povaze.

Proces administrace a úprava administrace u klinických skupin:

Administrace WJ IV COG u lidí s úzkostí byla mírně pozměněna podle obecných doporučení ke změně administrace (viz následující část) z důvodu přihlédnutí k férovosti při testování. Při administraci byl specifický poměrně velký kumulativní počet minut pauz, které respondenti potřebovali k pokračování v testování, ačkoli počet pauz byl malý: průměrně 22,0 min ($SD = 29,2$ min; rozpětí = 0–126 min). Kromě toho 23 % osob s úzkostí v souboru potřebovalo administraci rozdělit na dvě setkání.

Administrace WJ IV COG u lidí s depresí probíhala standardně bez modifikací. Ačkoli se klinická a kontrolní skupina nelišila v průměrném čase, byly pozorovány rozdíly. Klinická skupina s depresí vykazovala potřebu velkého množství přestávek a velkou variabilitu v délce trvání celého vyšetření, protože někteří nedokončili některé subtesty. Možnosti předčasného ukončení využili jen někteří respondenti z klinické skupiny, kteří byli následně ze studie vyřazeni ($n = 6$) a to z důvodů somatických obtíží, zhoršení aktuální nálady i celkového stavu či nadměrně zvýšenou mírou obav z neúspěchu a demotivace. Dále, 10 % klinických respondentů potřebovalo administraci rozdělit na dvě setkání.

Administrace WJ IV COG u lidí se schizofrenií probíhala standardně bez modifikací. Nebyl zjištěn rozdíl v naměřeném čase mezi klinickou skupinou se schizofrenií a kontrolní skupinou. Nicméně, 33 % klinických respondentů potřebovalo administraci rozdělit na dvě setkání. Bylo potřeba vždy zohlednit potenciální ohrožení výsledků účinkem psychofarmak a potenciálním propuknutím psychotické ataky v průběhu vyšetření.

Obecná doporučení k administraci u klinických skupin:

Při testování člověka z některé z klinických skupin doporučujeme zaměřit se na následující úpravu administrace. Standardní administrace WJ IV COG nedovoluje opakovat instrukce⁹¹. Avšak pokud bychom při testování klinických skupin u některých testů nezopakovali instrukce, budeme pravděpodobně měřit spíše úroveň koncentrace než kognitivních schopností (což je diskutabilní např. u testů, které měří pozornost nebo krátkodobou paměť). To je problematické zejména z toho důvodu, že snížená koncentrace pozornosti přímo souvisí s depresivní (Peckham, McHug, Otto, 2010), úzkostnou (Pergamin-Hight, 2015) i psychotickou (Fioravanti, Carlone, Vitale, Cinti a Clare, 2005) symptomatikou, a může tak intervenovat do procesu vyšetření a testové výsledky nepřiměřeně zkracovat.

Pacienti mohou zmeškat začátek instrukcí k úkolu nebo celé instrukce např. kvůli tomu, že se jejich pozornost obrátí dovnitř na vlastní negativní prožívání a vnímání zadání neproběhne. Pokud se tak stane, může to zvýšit tendenci vzdát celý test. Závěčné položky jsou vhodným nástrojem. Administrátor musí být schopen trpělivosti někdy na delší dobu potřebnou k zaznamenání instrukcí pacientem. Je potřeba se ujistit, že pacient instrukce slyšel a je připravený na testové položky.

Během administrace je potřeba být přijímající, nehodnotící, podporující, aktivně naslouchat a tvořit příjemnou a bezpečnou atmosféru a pracovní alianci. Přestože jde o schopnosti důležité při administraci testu jakémukoli klientovi, v případě pacientů (a zvláště pak hospitalizovaných) hrají ještě významnější úlohu. Administrátor by měl dát najevo, že je schopný pochopit těžkou životní situaci, v níž se pacient aktuálně nachází. Pacienti potřebují vědět, že administrátorovi na nich záleží, že jsou na stejné straně a směřují k dosažení společného cíle. Úsměv a příjemné vystupování administrátora jsou nápomocné faktory k vytvoření dobré pracovní aliance. Administrátor by měl podporovat pacienta, kdykoli cítí, že klesá jeho sebedůvěra a zvyšuje se jeho sebepodceňování, měl by ujistit pacienta, že výsledky z testování neovlivní jeho pobyt v zařízení (pokud je to pravda) a zmírnit tlak na výkon.

Pacienti při debriefingu po testování reportovali, že pro zachování sebekonceptu potřebují vědět, že je administrátor a priori vnímá jako schopné lidi. Mohou svůj potenciální špatný výkon skrýt za situační faktory. U řady pacientů byly patrné sebe-handicapující strategie, které v důsledku snižovaly pozorovaný výkon – je nutné poskytovat co nejméně motivů k takovémuto jednání. Když je možné (u některých testů) poskytovat zpětnou vazbu, lze toho využít ke zdůraznění všech správně zodpovězených položek a k demonstraci upřímné radosti z dobrého výkonu participanta, ve snaze změnit jeho obecné negativistické nastavení a zaměření pouze na ty nesprávně zodpovězené položky směrem k pozitivitě a naději. Pokud je stres způsoben potřebou odpovědět na položky správně, administrátor by měl zdůraznit, že ne všechny odpovědi musí být zodpovězeny a že obvykle lidé nemají plný počet bodů, a přesto mohou dosáhnout relativně dobrých výsledků v těchto výkonových testech.

Ve volném čase během pauzy, nebo když pacient potřebuje krizovou intervenci, lze poskytnout podpůrný nekonfrontační rozhovor a vést pacienta k uvědomění si principů emocionální regulace, např. když se začne vyskytovat vyhyčivé chování a únikové reakce. Někdy může být užitečným zdrojem zvládnutí testové situace také užití medikace, zejména když symptomy úzkosti jsou příliš zvýšeny. Další doporučení se vztahuje ke ztrátě motivace, únavě a problémům s udržením pozornosti. Pacienti se mohou snažit vyvarovat nepříjemným situacím, nicméně je důležité je laskavě ale konstantně podporovat v pokračování a v tom, aby se snažili a zkusili se znovu soustředit na další položku. Po skončení administrace je důležitá podpora a opět přijímající, vysvětlující debriefing, neboť test je dlouhý a také lidé bez klinické diagnózy z kontrolní skupiny na konci bývají vyčerpaní. Je vhodné ptát se na proces testování a získat zpětnou vazbu.

Specifická doporučení pro testování člověka s úzkostí, depresí a schizofrenií:

Kromě výše zmíněných obecných zásad, které platí pro všechny tři klinické skupiny, u každé klinické skupiny platí ještě specifické principy, které se v ostatních skupinách nevyskytly. V rámci skupiny pacientů s úzkostí je vhodné ještě před administrací redukovat strach ze ztráty kontroly tím, že poskytneme pacientovi s úzkostí co nejvíce informací ohledně celého procesu administrace, edukujeme jej o potenciálních symptomech, které mohou nastat v průběhu výkonové situace, o zvyšující se obtížnosti položek, o přibližné délce trvání testu. Důležité je nepřekvapovat pacienta a vysvětlovat krok po kroku. Strach ze ztráty kontroly, nízká tolerance k neurčitosti, strach ze selhání atd. mohou souviset s tendencí „být perfektní“ ve svém výkonu a tato tendence ho může paralyzovat v situaci, na kterou není připraven. Pacienti s úzkostí často mohou vyžadovat přesná čísla a přesný popis procedur, které se budou dít, ještě než se do samotného testování pustí, aby se na to mohli psychicky připravit. Díky této přípravě mohou subjektivně vnímat kontrolu nad situací.

Pro klinické skupiny deprese a schizofrenie dohromady platí doporučení zvážit aktuální stav testovaného člověka před začátkem administrace, protože vyšetření může způsobit zhoršení symptomů duševní nemoci, aby se předešlo potenciální dekompenzaci. Další doporučení se týká omezení zátěže testovaného člověka v podobě volby jen některých testů z WJ IV COG namísto administrace celé komplexní kognitivní baterie.

V rámci skupiny lidí s depresí se také objevovaly obavy před samotným testováním. Na rozdíl od úzkostné klinické skupiny však obavy přetrvávaly v průběhu celé administrace a po ukončení testu došlo namísto úlevy ještě ke zhoršení stavu u většiny respondentů (70 %). Lidé s depresí také výrazně častěji vyžadovali pauzu, v některých případech dokonce po každém testu.

V rámci skupiny lidí se schizofrenií se také objevovala potřeba zjednodušit instrukce, aby došlo k pochopení. Schizofrenní porucha však může zasahovat do standardních procesů chápání instrukcí a je potřeba věnovat mnohem větší úsilí a trpělivost vysvětlování zadání než u ostatních klinických skupin. Specificky se také objevovaly paranoidní pocity, spojené s potřebou testování důvěryhodnosti. Často a opakovaně se např. ptali na účel vyšetření, informace o zachování anonymity atd. Je také nezbytné přihlídnout k neustálému riziku propuknutí psychotického stavu během administrace.

Stručné výsledky jednotlivých diplomových prací

Podrobné informace jsou součástí dílčích diplomových prací, zde prezentujeme jen stručné výsledky, týkající se rozdílů klinických skupin a kontrolního vzorku.

Na vzorku pacientů s úzkostí byl spočítán bootstrapovaný ($n = 10.000$, BCa) párový t-test pro všech 18 subtestů a následně také pro všech devět širokých kognitivních schopností (CHC) a tři kompozitní skóry (GIA, BIA a Gf-Gc). Před analýzami byly všechny hodnoty převedeny do z-skórové metriky k získání průměru 0 a SD 1. Výsledky pro dílčí testy naznačují, že pacienti s úzkostí skórovali lépe než kontrolní skupina v testech Všeobecné znalosti (T8): $t(25) = 2,13$, $p < 0,05$, bootstrapovaný rozdíl průměrů = 0,51, $CI_{(95\%)} = [0,07; 0,98]$, Cohenovo $D_z = 0,42$; a v Opakování pseudoslov (T12): $t(25) = 2,33$, $p < 0,05$ bootstrapovaný rozdíl průměrů = 0,60, $CI_{(95\%)} = [0,10; 1,11]$, Cohenovo $D_z = .46$. Na druhou stranu pacienti s úzkostí skórovali hůře v testech Hledání čísel (T11): $t(25) = -2,24$, $p < 0,05$, bootstrapovaný rozdíl průměrů = -0,50, $CI_{(95\%)} = [-0,95; -0,05]$ Cohenovo $D_z = -0,44$; a Řazení názvů a čísel (T16): $t(25) = -2,09$, $p < 0,05$, bootstrapovaný rozdíl průměrů = -0,54, $CI_{(95\%)} = [-1,03; -0,04]$, Cohenovo $D_z = -0,41$. Všechny efekty jsou spíše slabší. Ve zbývajících testech rozdíly prokázány nebyly. Výsledky pro široké kognitivní schopnosti naznačují, že byl jediný pozorovaný rozdíl v CHC schopnosti Porozumění - Znalosti, kde pacienti s úzkostí dosahovali mírně lepších výsledků než kontrolní skupina: $t(25) = 2,17$, $p < 0,05$ bootstrapovaný rozdíl průměrů = 0,50, $CI_{(95\%)} = [0,03; 1,00]$, Cohenovo $D_z = .43$. Ve zbývajících širokých kognitivních schopnostech ani klastrových skórech rozdíly prokázány nebyly. Kromě toho byla spočtena jednoduchá bootstrapovaná ($n = 1.000$, BCa) lineární regresní analýza pouze na vzorku pacientů s úzkostí bez kontrolní skupiny ($n = 26$), kterým byl před vyšetřením administrován také měřicí nástroj Beck's Anxiety Inventory (BAI, Beck & Steer, 1988). Bylo zjišťováno, zda mohou být dílčí WJ-IV-COG subtesty predikovány mírou úzkostnosti před započítáním testování. Ukázalo se, že míra úzkostnosti může hrát roli v snížení výkonu u pozornostních testů Verbální pozornost (T3): $B = -0,02$, $p < 0,05$, $CI_{(95\%)} = [-0,04; -0,002]$ a Řazení názvů a čísel (T16): $B = -0,03$, $p < 0,01$, $CI_{(95\%)} = [-0,05; -0,02]$. Vztah mezi mírou úzkostnosti a ostatními testy nebyl prokázán. Pro přidruženou tematickou analýzu pozorování a výpovědí respondentů lze nahlédnout do diplomové práce (Klocek, 2019).

Na vzorku pacientů s depresí byl spočítán neparametrický Mann-Whitney U a párový Wilcoxon signed rank test pro několik vybraných dílčích testů. Zde reportujeme pouze výsledky Wilcoxonova testu. Analýzy byly počítány na hrubých skórech. Výsledky naznačují, že pacienti s depresí skórovali hůře než kontrolní skupina ve všech sledovaných testech a subtestech, konkrétně Číselné řady (T2): $W = 99,0$, $p < 0,01$, rozdíl průměrů = -3,50, Cohenovo $D = -0,52$; Hledání písmen (T4): $W = 52,0$, $p < 0,01$, rozdíl průměrů = -9,00, Cohenovo $D = -0,81$; Vypravování příběhů (T6): $W = 134$, $p < 0,05$, rozdíl průměrů = -7,00, Cohenovo $D = -0,41$; Vizualizace - Prostorové vztahy (T7A): $W = 106,0$, $p < 0,05$, rozdíl průměrů = -1,50, Cohenovo $D = -0,26$; Vizualizace - Rotace s kostkami (T7B): $W = 104,0$, $p < 0,05$, rozdíl průměrů = -3,00, Cohenovo $D = -0,40$; Formování konceptů (T9): $W = 23,0$, $p < 0,01$, rozdíl průměrů = -9,50, Cohenovo $D = -1,11$; Hledání čísel (T11): $W = 48,5$, $p < 0,05$, rozdíl průměrů = -12,00, Cohenovo $D = -0,81$; Audio-vizuální učení A (T13A): $W = 85,0$, $p < 0,01$, rozdíl průměrů = -7,00, Cohenovo $D = -0,67$; Audio-vizuální učení B (T13B): $W = 108,5$, $p < 0,01$, rozdíl průměrů = -10,50, Cohenovo $D = -0,49$; Audio-vizuální učení C (T13C): $W = 111,0$, $p < 0,05$, rozdíl průměrů = -11,00, Cohenovo $D = -0,35$; Rozpoznávání obrázků (T14): $W = 71,5$, $p < 0,01$, rozdíl průměrů = -2,00, Cohenovo $D = -0,53$; Analýza-syntéza (T15): $W = 121,0$, $p < 0,05$, rozdíl průměrů = -3,00, Cohenovo $D = -0,09$; Vyhledávání dvojic (T17): $W = 41,0$, $p < 0,01$, rozdíl průměrů = -28,00, Cohenovo $D = -0,95$. Kromě toho bylo zjišťováno, zda míra depresivity měřená pomocí Beck's Depression Inventory (BDI-II) souvisí s některým testem WJ-IV-COG pomocí

Spearmanova korelačního koeficientu. Byly nalezeny pouze slabší souvislosti mezi BDI-II a testem T13. Pro další informace lze nahlédnout do diplomové práce (Vítková, 2017).

Na vzorku pacientů se schizofrenií byl spočítán neparametrický Mann-Whitney U pro stejné testy a subtesty jako u pacientů s depresí. Analýzy byly počítány na hrubých skórech. Výsledky naznačují, že pacienti se schizofrenií skórovali hůře než kontrolní skupina ve všech sledovaných testech, konkrétně Číselné řady (T2): $U = 221$, $p < 0,01$, rozdíl průměrů = -5,00, Cohenovo $D = -0,92$; Hledání písmen (T4): $U = 108,5$, $p < 0,01$, rozdíl průměrů = -12,00, Cohenovo $D = -1,66$; Vypravování příběhů (T6): $U = 170,5$, $p < 0,01$, rozdíl průměrů = -15,00, Cohenovo $D = -1,24$; Vizualizace – Prostorové vztahy (T7A): $U = 191$, $p < 0,01$, rozdíl průměrů = -3,00, Cohenovo $D = -1,17$; Vizualizace – Rotace s kostkami (T7B): $U = 153$, $p < 0,01$, rozdíl průměrů = -5,00, Cohenovo $D = -1,29$; Formování konceptů (T9): $U = 120,5$, $p < 0,01$, rozdíl průměrů = -11,00, Cohenovo $D = -1,57$; Hledání čísel (T11): $U = 72,5$, $p < 0,01$, rozdíl průměrů = -17,00, Cohenovo $D = -1,95$; Audio-vizuální učení A (T13A): $U = 637,5$, $p < 0,05$, rozdíl průměrů = -6,00, Cohenovo $D = -0,86$; Audio-vizuální učení B (T13B): $U = 515$, $p < 0,01$, rozdíl průměrů = -14,00, Cohenovo $D = -1,01$; Audio-vizuální učení C (T13C): $U = 498,5$, $p < 0,01$, rozdíl průměrů = -15,00, Cohenovo $D = -1,03$; Rozpoznávání obrázků (T14): $U = 246$, $p < 0,01$, rozdíl průměrů = -2,00, Cohenovo $D = -0,91$; Analýza-syntéza (T15): $U = 188$, $p < 0,01$, rozdíl průměrů = -5,00, Cohenovo $D = -1,20$; Vyhledávání dvojic (T17): $U = 99,5$, $p < 0,01$, rozdíl průměrů = -39,00, Cohenovo $D = -1,78$. Pro další informace lze nahlédnout do diplomové práce (Schimová, 2017).

Celkově lze z předchozích výsledků vyčíst velké rozdíly mezi pacienty s depresí nebo schizofrenií a kontrolní skupinou a spíše nevýrazné rozdíly mezi pacienty s úzkostí a kontrolní skupinou. Kromě dalších možných důvodů se můžeme domnívat, že rozdíly ve skórování byly způsobeny buď typem sledované populace, nebo typem administrace. Výsledky mohou odkazovat na skutečné rozdíly v kognitivních schopnostech u klinických skupin a zdravé populace. Onemocnění depresí a schizofrenií může částečně snižovat kognici a onemocnění úzkostí může kognici v některých oblastech částečně zvyšovat a v jiných částečně snižovat. Pravděpodobnější však je, že právě dodržением standardního postupu administrace u klinických skupin s depresí a schizofrenií došlo k viditelně horšímu výsledku v měření kognitivních schopností a upravením administrace pro potřeby klinické skupiny s úzkostí se tyto rozdíly neprokázaly, naopak mohly být některé hodnoty uměle nadhodnoceny. Přikláníme se tedy spíše k tomu, že nebyly pozorovány závažné rozdíly v jednotlivých kognitivních schopnostech, pokud byl standardní postup administrace mírně upraven. Ke specifickým doporučením týkajícím se administrace by mělo být přihlédnuto při interpretaci výsledků. S přihlédnutím k doporučením a patřičnou úpravou administrace tedy lze použít testovou baterii Woodcock-Johnson´s IV COG u všech tří skupin a vyhodnocovat výsledky podle standardizačních norem. Výsledky je nutno interpretovat se zvýšenou obezřetností a zvažovat nekognitivní aspekty vyšetření, stejně jako kognitivní aspekty, které však přímo nesouvisí s právě testovanou schopností.

Literatura

- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, 131, 30–60.
- Ackerman, P. L., & Cianciolo, A. T. (2000). Cognitive, perceptual speed, and psychomotor determinants of individual differences during skill acquisition. *Journal of Experimental Psychology Applied*, 6, 259–290.
- Aldenderfer, M. S., & Blashfield, R. K. (1984). *Cluster analysis*. Thousand Oaks, CA: Sage Publications.
- American Association on Intellectual and Developmental Disabilities. (2010). *Intellectual disability: Definition, classification, and systems of support* (11th ed.). Washington, DC: Author.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2001). *Standardy pro pedagogické a psychologické testování*. Praha: Testcentrum.
- Andrich, D. (1978). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement*, 2, 581–594.
- Apel, K. (2011). What is orthographic knowledge? *Language, Speech, and Hearing Services in the Schools*, 42, 592–603.
- Arbuckle, J. L. (2012). *IBM SPSS AMOS 21 users guide*. Crawfordville, FL: Amos Development.
- Archibald, L. M. D., & Gathercole, S. E. (2006). Nonword repetition: A comparison of tests. *Journal of Speech, Language, and Hearing Research*, 49, 970–983.
- Arend, I., Colom, R., Botella, J., Contreras, M. J., Rubio, V., & Santacreu, J. (2003). Quantifying cognitive complexity: Evidence from a reasoning task. *Personality and Individual Differences*, 35, 659–669.
- Babcock, B., & Albano, A. D. (2012). Rasch scale stability in the presence of item parameter and trait drift. *Applied Psychological Measurement*, 36(7), 565–580.
- Baird, G., Slonims, V., Simonoff, E., & Dworzynski, K. (2011). Impairment in non-word repetition: A marker for language impairment or reading impairment? *Developmental Medicine and Child Neurology*, 53(8), 711–716.
- Barker, T. A., Torgesen, J. K., & Wagner, R. K. (1992). The role of orthographic processing skills on five different reading tasks. *Reading Research Quarterly*, 27, 335–345.
- Benson, J., & Hagtvet, K. A. (1996). The interplay between design, data analysis and theory in the measurement of coping. In M. Zeidner & N. Endler (Eds.), *Handbook of coping: Theory, research, applications* (pp. 83–106). Hoboken, NJ: Wiley.
- Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice*, 17(1), 10–22.
- Berninger, V. W. (1990). Multiple orthographic codes: Key to alternative instructional methodologies for developing the orthographic phonological connections underlying word identification. *School Psychology Review*, 19, 518–533.
- Berninger, V. W., Abbott, R., Thomson, J., Wagner, R., Swanson, H. L., Wijsman E., & Raskind, W. (2006). Modeling developmental phonological core deficits within a working memory architecture in children and adults with developmental dyslexia. *Scientific Studies in Reading*, 10, 165–198.

- Bond, T., & Fox, C. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum.
- Borgen, F. H., & Barnett, D. C. (1987). Applying cluster analysis in counseling psychology research. *Journal of Counseling Psychology*, 34(4), 456-468.
- Braden, J., & Niebling, B. (2012). Using the joint test standards to evaluate the validity evidence for intelligence tests. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 739-757). New York, NY: Guilford Press.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.
- Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, task switching, and working memory. *Developmental Neuropsychology*, 19, 273-293.
- Campbell, T., & Fiske, D. (1959). Convergent and discriminant validation by multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytical studies*. New York, NY: Cambridge University Press.
- Carroll, J. B. (1998). Human cognitive abilities: A critique. In J. J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 5-24). Mahwah, NJ: Lawrence Erlbaum.
- Carroll, J. B. (2003). The higher stratum structure of cognitive abilities: Current evidence supports g and about ten broad factors. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 5-22). New York, NY: Pergamon Press.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245-276.0020
- Chun, M. M., Golomb, J. D., & Turke-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology*, 62, 73-101.
- Cohen, A., Fiorello, C., & Farley, F. (2006). The cylindrical structure of the Wechsler Intelligence Scale for Children-IV: A retest of the Guttman model of intelligence. *Intelligence*, 24, 587-591.
- Conway, C. M., Pisoni, D. B., & Kronenberger, W. G. (2009). The importance of sound for cognitive sequencing abilities: The auditory scaffolding hypothesis. *Current Directions in Psychological Science*, 18(5), 275-279.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443-507). Washington, DC: American Council on Education.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Daniel, M. H. (1997, zima). The making of a test: Norming. *Assesment Information Exchange*. San Antonio, TX: Pearson.
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford Press.
- de Bree, E., Wijnen, F., & Gerrits, E. (2010). Non-word repetition in Dutch children at risk of dyslexia and children with SLI: Results of a follow-up study. *Dyslexia*, 16, 36-44.
- Dehn, M. J. (2008). *Working memory and academic learning: Assessment and intervention*. Hoboken, NJ: Wiley.
- de Jong, P. F. (1998). Working memory deficits of reading disabled children. *Journal of Experimental Child Psychology*, 170, 75-76.
- de Jong, P. F., Seveke, M. J., & van Veen, M. (2000). Phonological sensitivity and the acquisition of new words in children. *Journal of Experimental Child Psychology*, 76, 275-301.
- Derakshan, N., & Eysenck, M. W. (2009). Anxiety, processing efficiency, and cognitive performance: New developments from attentional control theory. *European Psychologist*, 14(2), 168-176.
- Duntelman, G. H. (1989). *Principal components analysis*. Thousand Oaks, CA: Sage Publications.
- Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and

phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research*, 57, 421–436.

Efron, B., & Tibshirani, R. (1993). *An introduction to bootstrap*. New York, NY: Chapman and Hall.

Ekstrom, R. B., French, J. W., & Harman, M. H. (1979). Cognitive factors: Their identification and replication. *Multivariate Behavioral Research Monographs*, 79(2), 3–84.

Elliott, C. D. (2007). *Differential Ability Scales* (2nd ed.) San Antonio, TX: Pearson.

Enders, C. K. (2006). A primer on the use of modern missing-data methods in psychosomatic medicine research. *Psychosomatic Medicine*, 68, 427–436.

Engle, R. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11(1), 19–23.

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling*, 6(1), 56–83.

Fioravanti, M., Carlone, O., Vitale, B., Cinti, M. E., & Clare, L. (2005). A meta-analysis of cognitive deficits in adults with a diagnosis of schizophrenia. *Neuropsychology review*, 15(2), 73–95. doi: 10.1007/s11065-005-6254-9

Flanagan, D. P., Alfonso, V. C., & Ortiz, S. O. (2012). The cross-battery assessment approach: An overview, historical perspective, and current directions. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 459–483). New York, NY: Guilford Press.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2007). *Essentials of cross-battery assessment* (2nd ed.). Hoboken, NJ: Wiley.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Hoboken, NJ.

Flanagan, D. P., Ortiz, S. O., Alfonso, V. C., & Mascolo, J. T. (2006). *The achievement test desk reference: A guide to learning disability identification* (2nd ed.). Hoboken, NJ: Wiley.

Gathercole, S. E. (2004). Working memory and learning during the school years. *Proceedings of the British Academy*, 125, 365–380.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, 27, 513–543.

Gathercole, S. E., & Alloway, T. P. (2008). Working memory and classroom learning. In S. K. Thurman & C. A. Fiorello (Eds.), *Applied cognitive research in K-3 classrooms* (pp. 17–40). New York, NY: Routledge.

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language*, 28, 200–213.

Gathercole, S. E., Lamont, E., & Alloway, T. P. (2006). Working memory in the classroom. In S. J. Pickering (Ed.), *Working memory and education* (pp. 220–241). Burlington, MA: Elsevier Press.

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology*, 28, 887–898.

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1994). *The Children's Test of Nonword Repetition: A test of phonological working memory*. *Memory*, 2, 103–127.

Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: Bridging selective attention and working memory. *Trends in Cognitive Science*, 16(2), 129–135.

Goldman, R., Fristoe, M., & Woodcock, R. W. (1974). *Goldman-Fristoe-Woodcock Auditory Skills Battery*. San Antonio, TX: Pearson.

Goldman-Rakic, P. S. (1992). Working memory and the mind. *Scientific American*, 267(3), 110–117.

Gorsuch, R. L., & Zachary, R. A. (1985). Continuous norming: Implication for the WIS-R. *Journal of Clinical Psychology*, 41(1), 86–94.

- Henry, D. B., Tolan, P. H., & Gorman-Smith, D. (2005). Cluster analysis in family psychology research. *Journal of Family Psychology*, 19(1), 121–132.
- Horn, J. L. (1965). *Fluid and crystallized intelligence* (Unpublished doctoral dissertation). University of Illinois, Urbana-Champaign.
- Horn, J. L. (1988). Thinking about human abilities. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate psychology* (2nd ed., pp. 645–865). New York, NY: Academic Press.
- Horn, J. L. (1989). Models for intelligence. In R. Linn (Ed.), *Intelligence: Measurement, theory, and public policy* (pp. 29–73). Urbana: University of Illinois Press.
- Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. McGrew, J. K. Werder, and R. W. Woodcock, *WJ-R technical manual* (pp. 197–232). Rolling Meadows, IL: Riverside.
- Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: *Gf-Gc* theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53–91). New York, NY: Guilford Press.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55.
- Jaffe, L. E. (2009). *Development, interpretation, and application of the W score and the relative proficiency index* (Woodcock-Johnson III Assessment Service Bulletin No. 11). Rolling Meadows, IL: Riverside.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.
- Jensen, A. R. (2011). The theory of intelligence and its measurement. *Intelligence*, 39, 171–177.
- Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 294–316). Thousand Oaks, CA: Sage Publications.
- Kaufman, A. S. (2009). *IQ testing 101*. New York, NY: Springer.
- Kaufman, A. S., & Lichtenberger, E. O. (2002). *Assessing adolescent and adult intelligence* (2nd ed.). Boston, MA: Allyn & Bacon.
- Keith, T. Z., & Reynolds, M. R. (2010). Cattell-Horn-Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools*, 47(7), 635–650.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling*. New York, NY: Guilford Press.
- Klocek, A. (2019). *Validation Study of Czech Version of Woodcock-Johnson's IV Test of Cognitive Abilities within Anxiety Disorder population and Non-clinical Population* (Diploma thesis). Brno. Retrieved from: https://is.muni.cz/th/qwlz5/Klocek_2018_diploma_thesis.pdf.
- Kolen, M. J., & Brennan, R. L. (2010). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York, NY: Springer.
- Konold, T. R., Glutting, J. J., McDermott, P. A., Kush, J. C., & Watkins, M. M. (1999). Structure and diagnostic benefits of a normative subtest taxonomy developed from the WISC-III standardization sample. *Journal of School Psychology*, 37(1), 29–48.
- Křen, M., Cvrček, V., Čapka, T., Čermáková, A., Hnátková, M., Chlumská, L., ... Zasina, A. (2015). *SYN2015: reprezentativní korpus psané češtiny*. Praha: Ústav Českého národního korpusu FF UK. Dostupné na www.korpus.cz.
- Lenhard, W. & Lenhard, A. (2019). cNORM: Continuous Norming. R package version 1.1.8. URL <https://CRAN.R-project.org/package=cNORM>
- Li, X., & Sireci, S. (2013). A new method for analyzing content validity data using multidimensional scaling. *Educational and Psychological Measurement*, 73(3), 365–385.
- Liberman, I. Y., Shankweiler, D., & Liberman, A. M. (1989). The alphabetic principle and learning to read. In D. Shankweiler & I. Y. Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle* (pp. 1–33). Ann Arbor: University of Michigan Press.
- Linacre, J. M. (2012). *WINSTEPS* (Version 3.74.0) [Computer software]. Chicago, IL: Winsteps.com.

- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, 3, 635–694.
- Lohman, D. F., & Lakin, J. (2011). Reasoning and intelligence. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (2nd ed., pp. 419–441). New York, NY: Cambridge University Press.
- Machálek, T. (2014). *KonText - aplikace pro práci s jazykovými korpusy*. Praha: FF UK. Dostupné na kontext.korpus.cz.
- Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence*, 7(2), 107–127.
- Mather, N., & Wendling, B. J. (2014a). Examiner's Manual. *Woodcock-Johnson IV Tests of Achievement*. Rolling Meadows, IL: Riverside.
- Mather, N., & Wendling, B. J. (2014b). Examiner's Manual. *Woodcock-Johnson IV Tests of Cognitive Abilities*. Rolling Meadows, IL: Riverside.
- Mather, N., & Wendling, B. J. (2014c). Examiner's Manual. *Woodcock-Johnson IV Tests of Oral Language*. Rolling Meadows, IL: Riverside.
- Mayringer, H., & Wimmer, H. (2000). Pseudoname learning by German-speaking children with dyslexia: Evidence for a phonological learning deficit. *Journal of Experimental Child Psychology*, 75, 116–133.
- McGrew, K. S. (1986). *Clinical interpretation of the Woodcock-Johnson Tests of Cognitive Ability*. Boston, MA: Allyn & Bacon.
- McGrew, K. S. (1994). *Clinical interpretation of the Woodcock-Johnson Tests of Cognitive Ability-Revised*. Boston, MA: Allyn & Bacon.
- McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151–180). New York, NY: Guilford Press.
- McGrew, K. S. (2005). The Cattell-Horn-Carroll (CHC) theory of cognitive abilities: Past, present, and future. In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 136–202). New York, NY: Guilford Press.
- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research [Editorial]. *Intelligence*, 37, 1–10.
- McGrew, K. S. (2012, September). *Implications of 20 years of CHC cognitive-achievement research: Back-to-the-future and beyond CHC*. Paper presented at the Richard Woodcock Institute, Tufts University, Medford, MA.
- McGrew, K. S., Dailey, D., & Schrank, F. A. (2007). *Woodcock-Johnson III/Woodcock-Johnson III Normative Update score differences: What the user can expect and why* (Woodcock-Johnson III Assessment Service Bulletin No. 9). Rolling Meadows, IL: Riverside.
- McGrew, K. S., & Flanagan, D. P. (1998). *The intelligence test desk reference (ITDR): Gf-Gc cross-battery assessment*. Boston, MA: Allyn & Bacon.
- McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). Technical Manual. *Woodcock-Johnson IV*. Rolling Meadows, IL: Riverside.
- McGrew, K. S., Schrank, F. A., & Woodcock, R. W. (2007). Technical Manual. *Woodcock-Johnson III Normative Update*. Rolling Meadows, IL: Riverside.
- McGrew, K. S., & Wendling, B. (2010). CHC cognitive-achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools*, 47(7), 651–675.
- McGrew, K. S., Werder, J. K., & Woodcock, R. W. (1991). Technical Manual. *Woodcock-Johnson Psycho-Educational Battery-Revised*. Rolling Meadows, IL: Riverside.
- McGrew, K. S., Woodcock, R. W. (2001). Technical Manual. *Woodcock-Johnson III*. Rolling Meadows, IL: Riverside.

- McGrew, K. S., & Wrightson, W. (1997). The calculation of new and improved WISC-III subtest reliability, uniqueness, and general factor characteristic information through the use of data smoothing procedures. *Psychology in the Schools*, 34, 181-195.
- Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-100). Washington, DC: American Council on Education.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses in performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749.
- Michas, I. C., & Henry, L. A. (1994). The link between phonological memory and vocabulary acquisition. *British Journal of Developmental Psychology*, 12, 147-164.
- Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. New York, NY: Holt, Rinehart & Winston.
- Mosier, C. I. (1943). On the reliability of a weighted composite. *Psychometrika*, 8, 161-168.
- Mulaik, S. A., James, L. R., Van Alstine, J., Bennet, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin*, 105(3), 430-445.
- Muñoz-Sandoval, A. F., Woodcock, R. W., McGrew, K. S., & Mather, N. (2007). *Batería III Woodcock-Muñoz Normative Update: Pruebas de habilidades cognitivas*. Rolling Meadows, IL: Riverside.
- Nation, K., & Hulme, C. (2010). Learning to read children's phonological skills: Evidence from latent variable longitudinal study of reading and nonword repetition. *Developmental Science*, 14(4), 649-659.
- Neisser, U. (1967). *Cognitive psychology*. New York, NY: Appleton-Century-Crofts.
- Nekula, M., Rusínová, Z., Grepl, M., & Karlík, P. (Eds.). (2012). *Příruční mluvnice češtiny*. Praha: Nakladatelství Lidové noviny.
- Nikolai, T., Štěpánková, H., Michalec, J., Bezdíček, O., Horáková, K., Marková, H., Růžička, E., & Kopeček, M. (2015). Testy verbální fluence, česká normativní studie pro osoby vyššího věku. *Česká a slovenská neurologie a neurochirurgie*, 111(3), 292-299.
- Nunnally, J. S. (1978). *Psychometric theories*. New York, NY: McGraw-Hill.
- Pala, K., & Všianský, J. (2000). *Slovník českých synonym*. Praha: Nakladatelství Lidové noviny.
- Palková, Z. (1994). *Fonetika a fonologie češtiny*. Praha: Karolinum.
- Peckham, A. D., McHugh, R. K. and Otto, M. W. (2010), A meta-analysis of the magnitude of biased attention in depression. *Depress. Anxiety*, 27: 1135-1142. doi: 10.1002/da.20755.
- Pergamin-Hight, L., Naim, R., Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., & Bar-Haim, Y. (2015). Content specificity of attention bias to threat in anxiety disorders: a meta-analysis. *Clinical psychology review*, 35, 10-18. doi: 10.1016/j.cpr.2014.10.005.
- Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience*, 35, 73-89.
- Pufpaff, L. A. (2009). A developmental continuum of phonological sensitivity skills. *Psychology in the Schools*, 46(7), 679-691.
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago, IL: University of Chicago Press. (Původní práce publikována 1960).
- Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. New York, NY: Routledge.
- Revelle, W. (2019). psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA. URL <https://CRAN.R-project.org/package=psych>.
- Roid, G. H. (2003a). *Stanford Binet Intelligence Scales* (5th ed.). Austin, TX: PRO-ED.
- Roid, G. H. (2003b). Technical Manual. *Stanford Binet Intelligence Scales* (5th ed.). Austin, TX: PRO-ED.

- Říčan, P., & Laciga, J. (2015). *Krátký inteligenční test*. Otrokovice: Propsyco, s.r.o.
- Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology*, 85, 112–118.
- Schimová, M. (2017). *Validizace české adaptace testu Woodcock Johnson IV COG u pacientů s diagnózou schizofrenie* (Diplomová práce). Brno. Dostupné z: https://is.muni.cz/th/mz454/DP_Schimova_nvyegzwe.pdf.
- Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99–144). New York, NY: Guilford Press.
- Schneider, W. J., & McGrew, K. S. (2018). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (p. 73–163). New York, NY: The Guilford Press.
- Schrank, F. A., Mather, N., & McGrew, K. S. (2014a). *Woodcock-Johnson IV Tests of Achievement*. Rolling Meadows, IL: Riverside.
- Schrank, F. A., Mather, N., & McGrew, K. S. (2014b). *Woodcock-Johnson IV Tests of Oral Language*. Rolling Meadows, IL: Riverside.
- Schrank, F. A., McGrew, K. S., & Mather, N. (2014a). *Woodcock-Johnson IV*. Rolling Meadows, IL: Riverside.
- Schrank, F. A., McGrew, K. S., & Mather, N. (2014b). *Woodcock-Johnson IV Tests of Cognitive Abilities*. Rolling Meadows, IL: Riverside.
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *Journal of Educational Research*, 99(6), 323–337.
- Siegel, L. S., & Ryan, E. B. (1989). The development of working memory in normally achieving and subtypes of learning disabled children. *Child Development*, 60, 973–980.
- Smith, Jr., E. V., & Smith, R. (Eds.). (2004). *Introduction to Rasch measurement*. Maple Grove, MN: JAM Press.
- Stahl, J. A., & Muckle, T. (2007). Investigating drift displacement in Rasch item calibrations. *Rasch Measurement Transactions*, 21(3), 1126–1127.
- Stankov, L. (2000). Structural extensions of a hierarchical view on human cognitive abilities. *Learning and Individual Differences*, 12(1), 35–51.
- Stankov, L. (2005). g factor: Issues of design and interpretation. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 279–293). Thousand Oaks, CA: Sage Publications.
- Süß, H.-M., & Beauducel, A. (2005). Faceted models of intelligence. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 313–332). London, United Kingdom: Sage Publications.
- Swanson, H. L. (1994). Short-term memory and working memory: Do both contribute to our understanding of academic achievement in children and adults with learning disabilities? *Journal of Learning Disabilities*, 27(1), 34–50.
- Swanson, H. L., Ashbaker, H. J., & Lee, C. (1996). Learning disabled readers' working memory as a function of processing demands. *Journal of Experimental Child Psychology*, 61, 242–275.
- Šebesta, K., Goláňová, H., Křen, M., & Procházka, P.: SCHOLA2010: *Korpus mluvené češtiny ve škole - přepisy nahrávek vyučovacích hodin na českých základních a středních školách*. Praha: Ústav Českého národního korpusu FF UK. Dostupné na www.korpus.cz.
- Štorková, P., Preiss, M., & Kopeček, M. (2004). Efekt nácviku testu verbální fluence a testování alternativní verze: Pilotní studie. *Psychiatrie*, 8(3), 186–189.
- Tucker-Drob, E. M., & Salthouse, T. A. (2009). Confirmatory factor analysis and multidimensional scaling for construct validation of cognitive abilities. *International Journal of Behavioral Development*, 33(3), 277–285.

- Urbánek, T., Denglerová, D., Širůček, J. (2011). *Psychometrika. Měření v psychologii*. Praha: Portál.
- VandenBos, G. (2007). *APA dictionary of psychology*. Washington, DC: American Psychological Association.
- Vitková, M. (2017). *Validizace české adaptace testu Woodcock-Johnson IV COG u klinických skupin* (Diplomová práce). Brno. Dostupné z: <https://is.muni.cz/th/s9v4q/DP-Vitkova.pdf>.
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Mass, H. L., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, 7(6), 632–638.
- Wagner, R. K., Torgesen, J. K., Laughon, P., Simmons, K., & Rashotte, C. A. (1993). The development of young readers' phonological processing abilities. *Journal of Educational Psychology*, 85, 1–20.
- Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). The development of reading-related phonological processing abilities: New evidence of bi-directional causality from a latent variable longitudinal study. *Developmental Psychology*, 30, 73–87.
- Wass, S. V., Scerif, G., & Johnson, M. H. (2012). Training attentional control and working memory—Is younger, better? *Developmental Review*, 32, 360–387.
- Wechsler, D. (1955). *Wechsler Adult Intelligence Scale*. San Antonio, TX: Pearson.
- Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of Intelligence* (3rd ed.). San Antonio, TX: Pearson.
- Wechsler, D. (2003). *Wechsler Intelligence Scale for Children* (4th ed.). San Antonio, TX: Pearson.
- Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (4th ed.). San Antonio, TX: Pearson.
- Wickham, H. & Bryan, J. (2019). readxl: Read Excel Files. R package version 1.3.1. URL <https://CRAN.R-project.org/package=readxl>
- Wilhelm, O. (2005). Measuring reasoning ability. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 373–392). London, United Kingdom: Sage Publications.
- Wolfe, E. W. (2004). Equating and item banking with the Rasch model. In E. V. Smith, Jr. & R. M. Smith (Eds.), *Introduction to Rasch measurement*. Maple Grove, MN: JAM Press.
- Wolfe, E. W., & Smith, Jr., E. V. (2007). Instrument development tools and activities for measure validation using Rasch models: Part II—Validation activities. *Journal of Applied Measurement*, 8(2), 204–234.
- Woodcock, R. W. (1978). *Development and standardization of the Woodcock-Johnson Psycho-Educational Battery*. Rolling Meadows, IL: Riverside.
- Woodcock, R. W. (1984). A response to some questions raised about the Woodcock-Johnson II. Efficacy of the aptitude clusters. *School Psychology Review*, 13, 355–362.
- Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment*, 8, 231–258.
- Woodcock, R. W. (1994). Norms. In R. J. Sternberg (Ed.), *The encyclopedia of intelligence*. New York, NY: Macmillan.
- Woodcock, R. W. (1999). What can Rasch-based scores convey about a person's test performance? In S. E. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement: What every psychologist and educator should know* (pp. 105–127). Mahwah, NJ: Lawrence Erlbaum.
- Woodcock, R. W. & Dahl, M. N. (1971). *A common scale for the measurement of person ability and test item difficulty* (AGS Paper No. 10). San Antonio, TX: Pearson.
- Woodcock, R. W., & Johnson, M. B. (1977). *Woodcock-Johnson Psycho-Educational Battery*. Rolling Meadows, IL: Riverside.
- Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Psycho-Educational Battery—Revised*. Rolling Meadows, IL: Riverside.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001, 2007). *Woodcock-Johnson III*. Rolling Meadows, IL: Riverside.

World Health Organization. (2016). *International Statistical Classification of Diseases and Related Health Problems, ICD10: Version 2016*. Retrieved from <http://apps.who.int/classifications/icd10/browse/2016/en>

Wright, B. D., & Douglas G. A. (1975). *Best test design and self-tailored testing* (MESA Memorandum No. 19). Retrieved from <http://www.rasch.org/memo19.pdf>

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. Chicago, IL: MESA Press.

Wright, B. D., & Stone, M. S. (1979). *Best test design: Rasch measurement*. Chicago, IL: MESA Press.

Definice Cattellovy-Hornovy-Carrollovy (CHC) teorie kognitivních schopností¹

Následující informace o širokých a úzkých schopnostech v rámci CHC teorie jsou výtahem z rozsáhlého narativně laděného popisu současné teorie CHC v:

Schneider, W. J. & McGrew, K.S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (3rd ed.), (pp. 99-144). New York, NY: Guilford Press².

Definice, které jsou zde uvedené, představují pouhou „kostru“ informací obsažených ve výše uvedené kapitole. Pro pochopení detailů čtenářům velmi doporučujeme citovanou kapitolu přečíst. Dodáváme rovněž, že – jak je popsáno a vysvětleno v Kapitolách 1 a 5 tohoto manuálu – řada navrhovaných změn, týkajících se alternativních kódů některých širokých a úzkých schopností (např. *Gwm* namísto *Gsm*) a jejich revidovaných definic, je již součástí tohoto textu. Tyto změny byly motivovány poznatky kognitivní neurovědy a analýzami strukturní validity testů WJ IV (viz Kapitola 5). Změny se týkají zejména současného pojetí konstruktů pracovní paměti (*Gwm*), kontroly pozornosti (*Gwm-AC*), rychlosti přístupu k slovní zásobě (*Glr-LA*) a paměti pro zvuky řeči (*Ga-UM*). Navržené změny jsou v textu zvýrazněny *kurzívou*.

Doménově nezávislé obecné kapacity

Fluidní inteligence³ (Fluid reasoning, *Gf*)

Fluidní usuzování je volní, nicméně flexibilní kontrolou pozornosti vedoucí k řešení aktuálních problémů, které nemohou být vyřešeny pouze díky spolehnutí se na dříve naučené zvyky, schémata a scénáře. Fluidní inteligence je multidimenzionálním konstruktem, všechny její dílčí fasety nicméně slouží společnému cíli: řešení doposud neznámých problémů. Fluidní usuzování je nejvíce patrné v abstraktních úvahách, které méně závisí na předešlém učení. Je nicméně také součástí řešení každodenních problémů. Fluidní uvažování je typicky uplatňováno v součinnosti se podpůrnými znalostmi a automatizovanými reakcemi. To znamená, že jeho aktivace, a to i na nejkratší moment, nastává ve

¹ Tato příloha pochází z původní americké verze Technického manuálu. Autorkou překladu je Edita Chvojková.

² Materiál sumarizovaný v této příloze byl v původní americké verzi TM adaptován a publikován se svolením Guilford Press ©.

³ Tento manuál si neklade za cíl závazně kodifikovat českou terminologii spojenou s CHC. V případě faktorů označovaných zkratkou *Gf* používáme termíny fluidní inteligence (pokud chceme zdůraznit, že jde o schopnost) nebo fluidní usuzování nebo fluidní uvažování (pokud chceme zdůraznit, že jde o proces).

chvíli, kdy jsou stávající návyky, scénáře a schémata nedostatečné vzhledem k nárokům nové situace. Fluidní inteligence je rovněž patrná v inferenčním vyvozování závěrů, tvorbě pojmů, klasifikaci neznámých podnětů, zobecňování starých řešení na nové problémy a kontexty, generování a potvrzování hypotéz, identifikaci relevantních podobností, rozdílů a vztahů mezi různými objekty a idejemi, vnímání relevantních důsledků nově nabytých znalostí a extrapolaci rozumných odhadů v nejasných situacích.

- Indukce (Induction; I) schopnost sledovat fenomén a odhalit principy jeho fungování či pravidla, která určují jeho chování.
- Obecné sekvenční usuzování (General sequential reasoning; RG): Schopnost logicky uvažovat na základě známých premis a principů. Tato schopnost je také nazývána deduktivní usuzování či aplikace pravidel.
- Kvantitativní usuzování (Quantitative reasoning; RQ) Schopnost uvažovat, ať už induktivně či deduktivně, s pomocí čísel, matematických vztahů a operátorů.

Krátkodobá pracovní paměť (Short-Term Working Memory; *Gwm*)

Krátkodobá pracovní paměť je schopnost kódovat, uchovávat a manipulovat informacemi v bezprostředním vědomí. *Gwm* odkazuje k individuálním rozdílům jednak v kapacitě (velikosti) primární paměti a jednak v efektivitě kontrolních mechanismů pozornosti, které manipulují informacemi v primární paměti.

- Rozsah paměti (Memory span; MS): Schopnost kódovat informace, udržovat je v primární paměti a okamžitě je reprodukovat v té stejné sekvenci, v níž byly prezentovány.
- Kapacita pracovní paměti (Working memory capacity; WM⁴): Schopnost zacílit ohnisko pozornosti na provádění relativně jednoduchých manipulací, kombinací a transformací informací v rámci primární paměti a současně se vyhýbat rozptylujícím podnětům a uplatňovat strategické a kontrolovaná vyhledávání informací ze sekundární paměti.
- *Kontrola pozornosti (Attentional control; AC): Schopnost soustředit se na podněty relevantní vzhledem k úkolu a ignorovat ty, které jsou nerelevantní. Schopnost regulovat cílevědomou zaměřenost a přímé kognitivní zpracování. Někdy také pojmenovávána jako „hledáček pozornosti“, ohnisko pozornosti, kontrola pozornosti, exekutivně kontrolovaná pozornost či exekutivní pozornost.*⁵

Dlouhodobá paměť a vybavení (Long-Term Storage and Retrieval; *Glr*)

Dlouhodobá paměť a vybavení je schopnost ukládat, upevňovat a vybavovat si informace po časových úsecích v řádu minut, hodin, dní a let. Krátkodobá paměť pracuje s informacemi, které byly zakódovány před několika sekundami a musí být vybaveny po dobu jejich aktivního udržení v pracovní paměti. Testy krátkodobé paměti často zahrnují informace uložené v té dlouhodobé. To, co *Gwm* od *Glr* odlišuje, je kontinuální úsilí o vědomé udržení této informace. *Glr* zahrnuje informace, které byly mimo bezprostřední uvědomění po dobu dostatečně dlouhou na to, aby mohl být obsah primární paměti kompletně nahrazen. V rámci *Glr* je nepřetržité udržení informací v primární paměti obtížné, ne-li nemožné.

Glr-Efektivita učení

Všechny úkoly na efektivitu učení musí prezentovat více informací, než může být udrženo v *Gwm*.

- Asociační paměť (Associative memory; MA): Schopnost pamatovat si dříve nesouvisějící informace na základě jejich párování.

⁴ Tento faktor byl dříve nazýván *pracovní paměť*. Nicméně, jak vysvětluje McGrew (2005), tento termín neodkazuje k proměnné chápané jako rozdíly mezi lidmi, ale spíše jako sada navzájem spřízněných kognitivních struktur. *Kapacita pracovní paměti* je oproti tomu proměnnou chápanou jako interindividuálními rozdíly, která je vlastností systému pracovní paměti jako celku.

⁵ Carroll (1993) identifikoval faktory, na něž se odkazoval jako na „schopnost všimati si (AC)“ a které popsal ve své kapitole „Různé oblasti schopností a osobnostních charakteristik“ (Miscellaneous Domains of Ability and Personal Characteristics). Carroll uvádí, že „jen málo faktorově-analytických studií bralo schopnosti pozornosti v potaz“ (s. 548). Navrhovaná definice pro *kontrolu pozornosti* (AC) je založena na současném kognitivně-neurovědním výzkumu (viz Chun, Golomb, & Turk-Brown, 2011; Derakshan & Eysenck, 2009; Engle, 2002; Gazzaley & Nobre, 2012; Petersen & Posner, 2012; Wass, Scerif, & Johnson, 2012). Je příjemnou náhodou, že kód Carrollova faktoru AC může zastupovat i *kontrolu pozornosti* (*attentional control*). Navrhujeme tedy, aby definice *kontroly pozornosti* navrhovaná zde, založená na současném neurovědě, dále užívala Carrollův kód „AC“ a reprezentovala právě kontrolu pozornosti.

- Smysluplná paměť (Meaningful memory; MM): Schopnost zapamatovat si vyprávění a jiné formy sémanticky souvisejících informací.
- Volné vybavení (Free recall memory, M6): Schopnost vybavit si seznamy v jakémkoli pořadí.

Glr-Fluence vybavení⁶

Glr-fluence vybavení reprezentuje rychlost a plynulost přístupu k informacím uloženým v dlouhodobé paměti.

Faktory fluence, které zahrnují produkci idejí:

- Ideální fluence (Ideational fluency; FI): Schopnost rychle vyprodukovat sérii idejí, slov či frází souvisejících se specifickou podmínkou či objektem. Jde především o kvantitu, nikoli kvalitu či originalitu odpovědí.
- Asociační fluence (Associational fluency; FA): Schopnost rychle vyprodukovat sérii nových či přínosných idejí souvisejících s konkrétním konceptem. Oproti Ideální fluenci (FI) je upřednostňována kvalita spíše než kvantita odpovědí.
- Expresivní fluence (Expressional fluency; FE): Schopnost rychle vymyslet různé způsoby, jak vyjádřit stejnou myšlenku.
- Senzitivita vůči problémům/fluence alternativních řešení (Sensitivity to problems, alternative solution fluency; SP): Schopnost rychle přijít s řadou alternativních řešení konkrétního praktického problému.
- Originalita/kreativita (Originality/creativity; FO): Schopnost rychle vyprodukovat originální, chytré a podnětné odpovědi (vyjádření, interpretace) na dané téma, situaci, úkol.

Faktory fluence, které zahrnují vybavení slov:

- Snadnost pojmenovávání (Naming facility, NA): Schopnost rychle označovat objekty jejich jmény. V současném výzkumu čtení se tato schopnost nazývá rychlé automatické pojmenovávání (rapid automatic naming; RAN) či rychlost lexikálního přístupu.
- Slovní fluence (Word fluency, FW): Schopnost rychle produkovat slova, která sdílí *stejnou fonologickou* (např. fluence vybavování slov v závislosti na nápodobě fonologického charakteru) či *sémantické vlastnosti* (např. fluence vybavování slov v závislosti na významově podmíněné reprezentaci). Zahrnuje rovněž schopnost rychlé produkce slov, která sdílí nesémantické charakteristiky (např. fluence vybavení slov začínajících písmenem T).
- Rychlost lexikálního přístupu (Speed of lexical access; LA): Schopnost vybavit si rychle a plynule slova z vlastní slovní zásoby; verbální efektivita či automaticnost lexikálního přístupu.

Figurální (obrazové) schopnosti fluence:

- Figurální fluence (Figural fluency; FF): Schopnost rychle nakreslit či načrtnout co možná nejvíce věcí (či detailů) v závislosti na prezentaci bezesmyslného vizuálního podnětu (např. sady různých vizuálních prvků). Kvantita je upřednostňována před kvalitou.
- Figurální flexibilita (FX): Schopnost rychle nakreslit různá řešení figurálních (obrazových) problémů.

Rychlost zpracování (Processing speed; Gs)

Rychlost zpracování je schopnost provádět jak jednoduché, tak komplexní kognitivní úlohy repetitivní povahy rychle a plynule. Považujeme ji za vesměs podružnou (v porovnání s Gf a Gc) s ohledem na predikci výkonu během učící se fáze získávání dovedností. Důležitým prediktorem naučeného výkonu se nicméně stává ve chvíli, jakmile již lidé vědí, jak úkol splnit. Jinými slovy, ve chvíli, kdy už lidé vědí, jak úkol zvládnout, stále se liší v rychlosti a plynulosti, s nimiž jej řeší. Například dva lidé mohou stejně přesně sečítat, nicméně jeden z nich si může matematická fakta vybavovat relativně snadno, zatímco jiný musí nad každou odpovědí přemýšlet o půl sekundy déle a počítat na prstech.

- Rychlost vnímání (Perceptual speed; P): Rychlost s jakou mohou být porovnány vizuální podněty z hlediska podobností nebo rozdílů. Podobně jako je Indukce (I) jádrem Gf, je percepční rychlost jádrem Gs. Výzkumná zjištění (Ackerman, Beier, & Boyle, 2005; Ackerman

⁶ Fluence vybavení je v poslední revizi CHC teorie oddělena od dlouhodobé paměti a učení. Je samostatným faktorem v druhé vrstvě (viz Schneider, Mc Grew; 2018).

& Cianciolo, 2000; McGrew, 2005) naznačují, že percepční rychlost může být schopnost na střední vrstvě (nacházející se mezi širokými a úzkými schopnostmi), definovaná čtyřmi úzkými podschopnostmi: (a) Rozpoznáváním vzorců (Pattern recognition; Ppr) – schopností rychle rozpoznávat jednoduché vizuální vzorce; (b) Skenováním (Scanning; Ps) – schopností skenovat, porovnávat a vyhledávat vizuální podněty; (c) Paměti (Memory; Pm) – schopností plnit úlohy vyžadující vizuální percepční rychlost, které kladou velké nároky na bezprostřední Gwm a (d) tzv. Komplexem (Complex; Pc) – schopnost plnit úkoly zaměřené na rozpoznávání vizuálních vzorců, které na respondenta kladou další kognitivní nároky, např. na prostorovou vizualizaci, odhad a interpolaci, a rovněž zvýšené nároky na rozsah paměti.

- Rychlost řešení testu (Rate-of-test-taking; R9): Rychlost a plynulost, s jakou jsou řešeny jednoduché kognitivní testy. Optikou teorie CHC byla tato definice zúžena na jednoduché testy, které nezahrnují vizuální srovnávání (a tedy se nepřekrývají s percepční rychlostí [P]) či mentální aritmetiku (takže se nepřekrývají s číselnými schopnostmi [N]).
- Číselné dovednosti (Number facility; N): Schopnost manipulovat s čísly v pracovní paměti a rychlost porovnávání číselných vzorců. Zahrnuje rychlost, s níž jsou správně prováděny základní aritmetické operace. Přestože tento faktor obsahuje vybavování matematických faktů, číselné dovednosti vyžadují rychlé provádění jakýchkoli jednoduchých kalkulací (např. odečítání čísla 3 od sloupce dvouciferných čísel). Součástí Číselných dovedností není porozumění nebo organizace matematických problémů a nejsou ani hlavní komponentou matematického/kvantitativního usuzování či vyšších matematických dovedností.
- Rychlost čtení (plynulost) (Reading speed [fluency]; RS): Rychlost čtení textu při zachování plného porozumění. Je rovněž uvedena u Grw.
- Rychlost psaní (plynulost) (Writing speed [fluency]; WS): Rychlost, s níž mohou být tvořeny nebo kopírovány slova nebo věty. Je rovněž součástí Grw a Gps.

Rychlost reakcí a rozhodování (Reaction and Decision Speed; Gt)

Rychlost reakcí a rozhodování se projevuje při provádění velmi jednoduchých rozhodnutí či úsudků, kdy jsou položky předkládány jedna po druhé. Primární užití škál zachycujících Gt nalézáme ve výzkumných studiích. Pro výzkumníky je Gt zajímavá, protože umožňuje poskytnout vhled do podstaty g a základních charakteristik mozku (např. neurální efektivity). Jedním ze zajímavých aspektů Gt je nejen to, že rychlejší reakční čas v těchto velmi jednoduchých úlohách souvisí s komplexitou usuzování, ale i to, že to platí i o větší konsistenci (tj. menší variabilitě) reakčního času.

- Jednoduchý reakční čas (Simple reaction time; R1): Reakční čas při vystavení jednomu konkrétnímu stimulu (vizuálnímu či auditivnímu). R1 je často dělen na fáze rozhodovacího času (DT; např. čas, který je nutný k rozhodnutí reagovat a začít přesouvat prst z výchozí pozice) a pohybového času (MT; tedy čas, který je potřebný k přesunutí prstu z výchozího tlačítka na jiné, které slouží k fyzickému zadání a záznamu konkrétní odpovědi).
- Výběrový reakční čas (Choice reaction time; R2): Reakční čas potřebný při provádění velmi jednoduché volby. Například tehdy, kdy respondent vidí dvě tlačítka a musí zmáčknout to z nich, které se rozsvítí.
- Rychlost sémantického zpracování (Semantic processing speed; R4): Reakční čas potřebný tehdy, kdy rozhodnutí vyžaduje jednoduché kódování a mentální manipulaci s obsahem podnětu.
- Rychlost mentálního porovnávání (Mental comparison speed; R7): Reakční čas ve chvíli, kdy musí být podněty porovnány s ohledem na konkrétní charakteristiku či atribut.
- Inspekční čas (Inspection time; IT): Rychlost, s níž mohou být zpozorovány rozdíly v podnětech.

Psychomotorické tempo (Psychomotor speed, Gps)

Psychomotorické tempo je rychlost a plynulost, s níž mohou být provedeny fyzické pohyby těla. Při učení se dovednostem je Gps schopností, která podmiňuje rozdíly ve výkonnosti poté, kdy lidé ze srovnatelné populace (např. manuální pracovníci ze stejné továrny) procvičovali jednoduchou dovednost po dostatečně dlouhou dobu.

- Rychlost pohybu končetin (Speed of limb movement; R3): Rychlost pohybu rukou a nohou. Rychlost je měřena po zahájení pohybu. Přesnost není důležitá.
- Rychlost psaní (plynulost) (Writing speed [fluency]; WS): Rychlost, jakou mohou být opisována slova. Je rovněž součástí *Grw* a *Gps*.
- Rychlost artikulace (Speed of articulation; PT): Schopnost rychle vykonat sérii artikulací řečovým svalstvem.
- Pohybový čas (Movement time; MT): Čas potřebný k vykonání fyzického pohybu konkrétní části těla (např. prstu), aby mohla být provedena konkrétní reakce. Aktuální výzkum naznačuje, že MT může být schopností střední úrovně (mezi úzkými a širokými schopnostmi), představující druhou fázi reakčního času, jak je měřen v různých elementárních kognitivních úlohách (ECT; elementary cognitive tasks). MT může rovněž měřit rychlost pohybu prstů, končetin, či pohybů, kterých se účastní více končetin, a rovněž rychlost vokální artikulace (diadochokinesis; řecký ekvivalent pro „po sobě jdoucí pohyby“). Je rovněž součástí *Gt*.

Systémy získaných znalostí

Porozumění-Znalosti (Comprehension-Knowledge; *Gc*)

Porozumění-Znalosti zachycují hloubku a šířku znalostí a schopností, které jsou v konkrétní kultuře považovány za hodnotné. Každá kultura upřednostňuje jisté konkrétní schopnosti a dovednosti před jinými. *Gc* reflektuje stupeň, do něž se člověk naučil prakticky použitelné znalosti a ovládl žádoucí dovednosti. Podle definice je tedy nemožné měřit *Gc* nezávisle na kultuře. *Gc* tedy pokrývá teoreticky širší záběr, než jaký je měřen jakoukoli existující kognitivní baterií.

- Obecné (verbální) informace (General [verbal] information; K0): Šířka a hloubka znalostí, které konkrétní kultura považuje za zásadní, praktické, či jiným způsobem hodnotné pro každého.
- Vývoj jazyka (Language development; LD): Obecné porozumění mluvenému jazyku na úrovni slov, idiomů a vět. Stejně jako považujeme Indukci (I) za jádro *Gf*, je Vývoj jazyka (LD) jádrem *Gc*. Přestože v Carrollově modelu se jedná o svébytnou úzkou schopnost, popis jeho analýz jasně napovídá, že Vývoj jazyka je kategorií ležící mezi *Gc* a specifickějšími s jazykem spojenými schopnostmi, jakými jsou např. Slovní zásoba (VL), Gramatická senzitivita (MY) a Poslechové schopnosti (LS). Vývoj jazyka se zdá být zastřešujícím pojmem pro všechny jazykové schopnosti pracující ve vzájemné součinnosti.
- Slovní zásoba (Lexical knowledge; VL): Znalost definic slov a pojmů, které jsou za nimi. Zatímco Vývoj jazyka (LD) se týká spíše porozumění slovům v kontextu, Znalost lexika (VL) spíše porozumění definicím izolovaných slov.
- Poslechové schopnosti (Listening ability; LS): Schopnosti porozumět mluvené řeči. Testy poslechových schopností typicky používají jednoduchou slovní zásobu, ale postupně k poslechu prezentují stále komplexnější skladbu (syntax) nebo čím dál delší pasáže řeči.
- Komunikační schopnost (Communication ability; CM): Schopnost užívat řeč pro srozumitelnou komunikaci vlastních myšlenek. Tato schopnost je srovnatelná s Poslechovými schopnostmi (LS), je nicméně produktivní (expresivní), spíše než receptivní.
- Gramatická senzitivita (Grammatical sensitivity; MY): Uvědomění si formálních pravidel gramatiky a morfologie slov v mluvené řeči. Tento faktor se liší od Užití angličtiny (EU) tím, že se manifestuje na úrovni mluveného a nikoli psaného jazyka, a tím, že reflektuje uvědomění si gramatických pravidel spíše než jejich korektní užití.

Doménově specifické znalosti (Domain-Specific Knowledge; *Gkn*)

Doménově specifické znalosti pokrývají hloubku, šířku a míru ovládnutí specializovaných znalostí (znalostí, u nichž nepředpokládáme, že by je měli všichni lidé ve společnosti). Specializované znalosti jsou typicky získávány v průběhu kariéry, prostřednictvím koníčků či dlouhodobého zájmu (např. náboženského či sportovního).

- Zběhlost v cizím jazyce (Foreign language proficiency; KL): Je podobná Jazykovému vývoji (LD), jde však o zběhlost v cizím jazyce. Od Schopností v cizím jazyce (Foreign language aptitude⁷, LA) se naopak odlišuje tím, že vyjadřuje reálnou zběhlost, a ne pouze potenciál či nadání. Můžeme předpokládat, že mnoho lidí s vysokou Zběhlostí v cizím jazyce má rovněž vysoké Schopnosti v cizím jazyce (LA), nicméně ne všichni lidé s vysokými Schopnostmi v cizím jazyce (LA) dosáhli v užívání daného jazyka mistrovství. Dříve byla tato schopnost klasifikována jako jeden z aspektů Gc. Nicméně od té doby, kdy byly do CHC zařazeny Doménově specifické znalosti (Gkn), je zjevné, že klasifikace specializované znalosti konkrétního jazyka by měla být změněna. Přestože Znalost angličtiny jako druhého jazyka (KE) byla dříve zařazena jako oddělená schopnost v rámci Gkn, nyní je jasné, že jde o specifický případ obecnější schopnosti Zběhlost v cizím jazyce (KL). Povšimněme si, že tento faktor je neobvyklý, protože není jeden – existuje faktor Zběhlosti v cizím jazyce odlišný pro každý jazyk.
- Znalost znakování (Knowledge of signing; KF): Znalost znakování konkrétních písmen a znakové řeči jako takové (např. Amerického znakového jazyka).
- Schopnost odezírání (Skill in lip reading; LP): Kompetence porozumět komunikaci ostatních pozorováním pohybů jejich úst a jejich výrazů.
- Geografické znalosti (Geography achievement; A5): Šíře zeměpisných poznatků (např. o hlavních městech států).
- Obecná vědecká informovanost (General science information; K1): Rozsah vědeckých poznatků z oblasti přírodních věd (např. biologie, fyziky, techniky, mechaniky, elektrotechniky).
- Kulturní znalosti (Knowledge of culture; K2): Rozsah znalostí z humanitních oborů (např. filozofie, religionistiky, historie, literatury, hudby a umění).
- Mechanické znalosti (Mechanical knowledge; MK): Znalosti o funkci, terminologii a obsluze běžných nástrojů, strojů a vybavení. Existuje mnoho testů mechanických znalostí a usuzování používaných pro účely personálního výběru (např. Baterie pro testy schopností ozbrojených sil; *Armed Services Vocational Aptitude Battery* [ASVAB], Weisenův test mechanických schopností; *Weisen Test of Mechanical Aptitude* [WTMA]).
- Znalosti obsahů chování (Knowledge of behavioral content; BC): Znalosti či sensitivita vůči neverbálním systémům lidské komunikace/interakce (např. obličejové výrazy a gesta). Výzkum v oblasti emoční inteligence (EI) je velmi rozsáhlý, není však doposud jasné, které z konstruktů EI by měly být do teorie CHC zařazeny. Jelikož teorie CHC se soustředí spíše na schopnosti než na osobnost, měly by být schopnosti, které jsou její součástí, měřitelné testy, v nichž existují správné a nesprávné odpovědi (případně mají rychlostní složku).

Čtení a psaní (Reading and Writing; Grw)

Čtení a psaní postihuje hloubku a šířku znalostí a dovedností souvisejících s psaným jazykem. Lidé s vysokým Grw čtou bez většího úsilí a píšou rovněž jen s malými obtížemi. Pokud je Grw dostatečně vysoké, čtení a psaní poskytují náhled na úroveň jazykového vývoje konkrétního člověka. Jakékoli obtíže, které může mít s porozuměním textu či efektivní komunikací, jsou v takových případech nejspíše funkcí Gc nebo Gkn. Naopak u osob s nízkým Grw nemusí být vysoké jazykové schopnosti dobře patrné z výkonu při čtení či psaní. Přestože čtení a psaní jsou zcela evidentně rozdílné aktivity, obě jsou podmíněny společnými zdroji, jejichž individuální rozdíly v dovednostech čtení a psaní obě aktivity neumožňují jasně odlišit. Zdá se, že schopnost společná všem dovednostem z oblasti čtení vzájemně propojuje i dovednosti z oblasti psaní.

- Porozumění verbálnímu (psanému) jazyku (Verbal [print] language comprehension; V): Obecný vývoj či porozumění slovům, větám a odstavcům v rodném jazyce, tak jak je měřen testy slovní zásoby a čtení s porozuměním. Nezahrnuje psaní, poslech, či porozumění mluveným instrukcím.
- Dekódování při čtení (Reading decoding; RD): Schopnost rozpoznat slova v textu. Tato schopnost je typicky zjišťována ústními testy čtení se slovy uspořádanými ve vzestupném pořadí dle obtížnosti. Testy mohou sestávat z foneticky pravidelných slov (tedy těch, které zní stejně, jako jsou psány; např. vana, kapka), foneticky nepravidelných slov (slov, která nezní stejně, jako

⁷ V tomto překladu nerozlišujeme anglické termíny „ability“ a „aptitude“, které oba překládáme českým slovem „schopnost“.

je píšeme; např. dcera, tchyně), či foneticky pravidelných pseudoslov (nesmyslných slov, která však podléhají pravidlům slovní reprodukce textu v daném jazyce, např. „latyš“ nebo „mortyka“).

- Čtení s porozuměním (Reading comprehension; RC): Schopnost porozumět psanému sdělení. Čtení s porozuměním bývá měřeno celou řadou způsobů.
- Rychlost čtení (Reading speed; RS): Tempo, jímž je člověk schopen číst souvislé sdělení s plným porozuměním. Rychlost čtení je klasifikována jako smíšená míra G_s (kognitivní rychlosti) a G_{rw} v hierarchickém rychlostním modelu.
- Schopnost hláskování (Spelling ability; SG): Schopnost hláskovat slova. Tento faktor je typicky měřen tradičními testy hláskování psaného textu. Nicméně, stejně jako v případě Dekódování při čtení (RD) může být rovněž měřen testy hláskování obsahujícími foneticky pravidelná pseudoslova (např. „latyš“). Je na místě poznamenat, že Carroll (1993) považoval tento faktor za nedostatečně definovaný a vhodný pro další výzkum.
- Užití angličtiny (English usage; EU): Znalost mechanismů psaní (např. psaní velkých písmen, interpunkce a korektního použití slov).
- Schopnost psaní (Writing ability; WA): Schopnost použít text k jasné komunikaci myšlenek.
- Rychlost psaní (Writing speed; WS): Schopnost rychle kopírovat či generovat text. Úkoly zjišťující Rychlost psaní jsou považovány jak za míru G_{rw} , tak G_{ps} (širšího psychomotorického tempa) v hierarchické rychlostní taxonomii.

Kvantitativní znalosti (Quantitative knowledge; G_q)

Kvantitativní znalosti pokrývají hloubku a šířku znalostí spojených s matematikou. G_q se liší od kvantitativního usuzování (RQ ; fasety G_f), a to stejným způsobem jako se G_c liší od nekvantitativních aspektů G_f . Skládá se z nabytých znalostí o matematice, jako je znalost matematických symbolů (např. $\int$, π , Σ , ∞ , $\neq$, $\leq$, $+$, $-$, $\times$, $\div$, $\sqrt{}$ a mnoha dalších), operací (např. sčítání/odčítání, násobení/dělení, exponenciální funkce/ n -té odmocniny, faktoriály, negace a mnoha dalších), výpočetních postupů (např. dělení vícetřídních čísel, krácení zlomků, kvadratických rovnic a mnoha dalších) a jiných s matematikou spřízněných dovedností (např. použití kalkulačky, matematického softwaru a dalších matematických pomůcek).

- Matematické znalosti (Mathematical knowledge; KM): Rozsah obecných znalostí o matematice. Tento faktor se netýká výkonu v úlohách zaměřených na matematické operace ani řešení matematických problémů. Naopak je spíše faktorem zjišťujícím „co“, nikoli „jak“ (např. Co znamená ∞ ? Co je Pythagorova věta?).
- Výkon v matematice (Mathematical achievement; A3): Měřený (testy zjišťovaný) matematický výkon.

Senzorické/Motoricky podmíněné schopnosti (Sensory/Motor-Linked Abilities)

Vizuální zpracování (Visual processing, G_v)

Vizuální zpracování je schopností, která využívá simulovaných mentálních představ (často ve spojení s právě vnímanými obrazy) k řešení problémů. Jakmile skrz oči projde vizuální informace, zrakové systémy mozku automaticky provedou velké množství výpočtů na nižší úrovni zpracování (např. detekci okrajů, vnímání světla a tmy, rozlišení barev, detekci pohybu atd.). Výsledky těchto operací nižší úrovně jsou následně zpracovány různými organizačně vyššími procesory, aby z nich byly vyvozeny komplexní aspekty vizuálního obrazu (např. rozpoznání objektu, konstrukce modelu prostorové konfigurace, predikce pohybu atd.).

- Vizualizace (Visualization; Vz): Schopnost vnímat komplexní vzorce a mentálně simulovat, jak by mohly vypadat po transformaci (např. rotované, zmenšené či zvětšené, částečně zakryté atd.). Stejně jako je Indukce (I) centrální schopností G_f a Jazykový vývoj (LD) hlavní komponentou G_c , je tato schopnost jádrem G_v .

- Rychlostní rotace (prostorové vztahy; Speeded rotation [spatial relations]; SR): Schopnost rychle řešit problémy uplatněním mentální rotace jednoduchých obrázků. Tato schopnost je podobná Vizualizaci (Vz), protože zahrnuje otáčení mentálních obrazů, je ale rozdílná, protože vyžaduje *rychlost*, díky níž lze mentální rotační úkoly provést. Úlohy s rychlostními rotacemi obvykle využívají spíše jednoduché obrázky.
- Rychlost uzavření (Closure speed; CS): Schopnost rychle identifikovat povědomý, smysluplný vizuální objekt z nekompletního (např. vágního, částečně zakrytého, nepropojeného) vizuálního stimulu bez předchozí znalosti, o jaký objekt se jedná. Tato schopnost je občas nazývána rovněž Vnímání tvaru (*Gestalt perception*), protože od respondenta vyžaduje „doplnění“ neviděných či chybějících částí obrazu potřebné k vizualizaci jediného vjemu.
- Flexibilita uzavření (Flexibility of closure; CF): Schopnost identifikovat vizuální figuru či vzorec vetkaný do komplexního rozptýlujícího či maskujícího vizuálního vzorce či oblasti na základě předchozí znalosti daného vzorce.
- Vizuální paměť (Visual memory; MV): Schopnost zapamatovat si komplexní obrazy v krátkém časovém limitu (méně než 30 sekund). Úlohy definující tento faktor zahrnují expozici komplexních obrazů a jejich následnou identifikaci po odebrání podnětového materiálu.
- Prostorové skenování (Spatial scanning; SS): Schopnost rychle a přesně prozkoumat (vizuálně explorovat) široké či komplikované prostorové pole nebo vzorec a (a) identifikovat konkrétní cílovou konfiguraci, nebo (b) identifikovat cestu polem za účelem určení cílového bodu. Není jasné, zda je tato schopnost spojena s komplexními realistickými širokopásmovými navigačními schopnostmi.
- Sériová percepční integrace (Serial perceptual integration; PI): Schopnost rozpoznat objekt poté, kdy jsou v rychlém sledu prezentovány pouze jeho části.
- Odhad délky (Length estimation; LE): Schopnost vizuálně odhadovat délky objektů.
- Percepční iluze (Perceptual illusions; IL): Schopnost nenechat se oklamat zrakovými klamy.
- Percepční alternace (Perceptual alternations; PN): Konzistence v míře alternací mezi různými vizuálními percepce.
- Představivost (Imagery; IM): Schopnost mentálně si představovat živé obrazy. Detailní studie imaginace na úrovni mozku naznačují, že vizuální prostorová představivost nemusí být jedinou schopností, ale že vizualizace umístění v prostoru a mentální transformace této prostorové konfigurace jsou závislé na rozličných neurálních sítích. Na základě výzkumu je možné se domnívat, že transformační procesy a paměť na umístění jsou funkčně nezávislé substrukтуры. Byla také navržena dichotomie imaginace objektů a imaginace prostoru, stejně jako možnost rozlišovat kvalitu a rychlost schopnosti představivosti.

Auditivní zpracování (Auditory processing; Ga)

Auditivní zpracování je schopností detekovat a zpracovat smysluplné neverbální zvukové informace. Tato definice může být trochu matoucí, protože nemáme dobře rozvinutý slovník pro komunikaci o zvucích mimo mluvené řeči či hudby. *Ga* zahrnuje obě tyto domény, ale ještě mnohem víc. Existují dvě chybná pojetí spojovaná s *Ga*. Zaprvé, přestože *Ga* závisí na senzoričtém vstupu, nejde o senzoričtý vstup jako takový. *Ga* je tím, co mozek provede se senzoričtí informací z ucha, někdy i dlouho poté, co jsme zvuk uslyšeli. Druhý omyl považuje *Ga* za porozumění mluvenému jazyku. Je pravdou, že jeden z aspektů *Ga* (rozbor řečových zvuků či fonetické kódování [PC]) je ve vztahu s porozuměním mluvené řeči, jde ale spíše o prekurzor porozumění, nikoli o porozumění samotné.

- Fonetické kódování (Phonetic coding; PC): Schopnost slyšet různé fonémy odlišně. Tato schopnost je někdy označována jako fonologické zpracování a fonologické uvědomění. Lidé se slabým fonetickým kódováním mají obtíže slyšet vnitřní strukturu zvuků ve slovech.
- Rozlišování zvukových vzorců (Speech sound discrimination; US): Schopnost detekovat a rozlišovat rozdíly ve zvucích řeči (jiných než konkrétních fonémech) za podmínek minimálního rozptýlení nebo zkreslení. Slabá schopnost diskriminace zvukových vzorců může způsobit obtíže ve schopnosti rozlišovat variabilitu v tónu, tónu, tónu a výšce řeči.

- Odolnost vůči zkreslení auditivního stimulu (Resistance to auditory stimulus distortion; UR): Schopnost slyšet slova správně i za přítomnosti zkreslení či výrazného hluku v pozadí.
- Paměť na zvukové vzorce (Memory for sound patterns; UM): Schopnost podržet (na krátkodobé bázi) auditivní *kódy jako tóny, tónové vzorce či řečové zvuky*.
- Udržování a rozpoznávání rytmu (Maintaining and judging rhythm; U8): Schopnost rozpoznat a udržet hudební rytmus. To může být jedním z aspektů paměti na zvukové vzorce (UM), protože je zde jasně zahrnuta krátkodobá paměť. Je nicméně pravděpodobné, že za rozlišováním rytmu stojí ještě další specifická komponenta, která si žádá odlišení.
- Hudební rozlišování a úsudek (Musical discrimination and judgement; U1 a U9): Schopnost rozlišit a posoudit tónové vzorce v hudbě s ohledem na melodické, harmonické a expresivní aspekty (frázování, tempo, harmonickou složitost, variaci intenzity).
- Absolutní sluch (Absolute pitch; UP): Schopnost perfektně identifikovat výšky tónů. (Neodpustíme si drobnou historickou poznámku, John Carroll měl absolutní sluch.)
- Zvuková lokalizace (Sound localization; UL): Schopnost lokalizovat slyšené zvuky v prostoru.

Olfaktorické schopnosti (Olfactory abilities, *Go*)

Olfaktorické schopnosti umožňují zachytit a zpracovat smysluplné informace z pachů. *Go* neodkazuje k citlivosti olfaktorického systému, ale k jakékoli kognici člověka, jakou je nos schopen vyslat dále. Doména *Go* pravděpodobně obsahuje mnohem více úzkých schopností, než je zahrnuto v současném CHC modelu. I velmi zběžná rešerše současného výzkumu *Go* a příbuzných oblastí odhalí zmínky o takových schopnostech jako je olfaktorická paměť, epizodická pachová paměť, olfaktorická senzitivita, pachově-specifické schopnosti, identifikace a detekce pachů, pojmenování pachů a pachová představivost, a tím výčet zdaleka nekončí.

- Olfaktorická paměť (Olfactory memory; OM): Schopnost rozpoznávat již dříve zaznamenané rozlišitelné pachy. OM hraje roli v čichově podmíněné zkušenosti, v jejímž rámci cítíme konkrétní pach či vůni, a následně jsme zaplaveni živými vzpomínkami na naše poslední setkání s tímto vjemem. Lidé mají tendenci uchovávat si vzpomínky na rozlišitelné pachy déle, než je tomu v případě jiných typů paměti.

Taktilní schopnosti (Tactile abilities, *Gh*)

Taktilní schopnosti zahrnují detekci a zpracování smysluplných informací haptických (dotykových) vjemů. *Gh* nepostihují individuální citlivost dotyků, nýbrž individuální kognice spojené s taktilními vjemy. Protože tato schopnost není zatím dobře definována a zcela pochopena, je velmi obtížné popisovat ji z pozice autority. Doména může obsahovat takové schopnosti jako Taktilní vizualizaci (identifikaci objektu na základě ohmatání), taktilní lokalizaci (např. rozpoznání místa dotyku na vlastním těle), taktilní paměť (např. vzpomínky na místa minulých dotyků), znalost textur (pojmenování povrchů a látek na základě hmatu) a mnoho dalších. V rámci *Gh* doposud nebyla podpořena existence žádných úzkých faktorů kognitivních schopností. Taktilní senzitivita (TS), schopnost senzorické přesnosti, odkazuje na schopnost přesně rozlišovat mezi haptickými vjemy (např. ve chvíli, kdy sevřeme dvě raménka svíracího měřítka současně na kůži, a osoba je vnímá jako jeden jediný bod, pokud jsou nedaleko od sebe – někteří lidé jsou v tomto ohledu schopni přesnějšího rozlišování než jiní).

Kinestetické schopnosti (Kinesthetic abilities; *Gk*)

Kinestetické schopnosti souvisí s detekcí a zpracováním smysluplných informací založených na vjemech z proprioreceptorů. Propriorecepce odkazuje ke schopnosti detekovat pozice a pohyby končetin prostřednictvím *proprioreceptorů* (senzorických orgánů ve svalcích a šlachách, které zaznamenávají natahování a smršťování). *Gk* nepostihuje samotnou senzitivitu propriorecepce, ale individuální kognici proprioreceptivních vjemů. V rámci *Gk* neznáme žádné úzké faktory kognitivních schopností, jejichž existence by byla dobře podpořena; nicméně určité studie naznačují, že kinestetická senzitivita (KS), schopnost senzorické přesnosti, je zodpovědná za rozlišování mezi proprioreceptivními vjemy (např. zda a jak moc bylo pohnuto s končetinou).

Psychomotorické schopnosti (Psychomotor abilities, *Gp*)

Jedná se o schopnosti vykonávat fyzické motorické pohyby těla (např. pohyby prstů, rukou, nohou) s přesností, koordinací nebo silou.

- Míření (Aiming; *A1*): Schopnost přesně a plynule vykonat sekvenci koordinovaných pohybů ruka-oko, a to konkrétně s potřebou zacílení.
- Manuální zručnost (Manual dexterity; *P1*): Schopnost přesně vykonávat koordinované pohyby ruky nebo současně ruky a paže.
- Prstová zručnost (Finger dexterity; *P2*): Schopnost provést přesně koordinované pohyby prstů (s manipulací s objekty nebo bez ní).
- Statická síla (Static strenght; *P3*): Schopnost vyvinout fyzickou sílu za účelem pohnout (zatlačit, zvednout, zatáhnout) relativně těžkým či imobilním objektem.
- Hrubá tělesná rovnováha (Gross body ekvilibrium; *P4*): Schopnost udržet tělo ve vzpřímené pozici v prostoru nebo znovu získat rovnováhu po jejím narušení.
- Koordinace více končetin (Multilimb coordination; *P6*): Schopnost činit rychlé, specifické či izolované motorické pohyby rukama či nohama.
- Koordinace paže-ruka (Arm-hand steadiness; *P7*): Schopnost precizně a dovedně koordinovat pozici paže a ruky v prostoru.
- Přesnost kontroly (Control precision; *P8*): Schopnost vyvinout přesnou kontrolu pohybu svalů, typicky v odpovědi na zpětnou vazbu z prostředí (např. změny v rychlosti a pozici manipulovaného objektu).

Příloha B

Souhrnné popisné statistiky a údaje o reliabilitě jednotlivých testů kognitivních schopností Woodcock-Johnson® IV

Test	Statistika	VĚK				
		5;0 – 6;11	7;0 – 8;11	9;0 – 10;11	11;0 – 12;11	13;0 – 14;11
Standardní baterie						
Test 1: Slovník	n	64	57	48	62	60
	M	472,7	491,9	506,2	511,1	517,7
	SD	20,2	17,2	13,6	11,9	11,9
	r _{CC}	0,979	0,977	0,969	0,960	0,960
	SE _m (W)	8,4	6,5	5,8	5,7	5,6
	SE _m (SS)	8,2	6,3	5,6	5,5	5,5
Test 2: Číselné řady	n	64	57	48	62	60
	M	433,9	487,8	511,6	521,5	528,4
	SD	39,6	25,5	18,8	19,9	16,0
	r _{CC}	0,994	0,990	0,983	0,985	0,978
	SE _m (W)	9,2	6,6	6,1	6,0	5,7
	SE _m (SS)	6,0	4,4	4,2	4,2	4,0
Test 3: Verbální pozornost	n	64	57	48	62	60
	M	466,1	487,2	504,6	512,9	517,5
	SD	24,9	18,7	12,3	11,9	11,4
	r _{CC}	0,986	0,981	0,962	0,961	0,958
	SE _m (W)	7,9	6,3	5,7	5,5	5,4
	SE _m (SS)	6,3	5,2	5,1	5,1	5,1
Test 4: Hledání písmen	n	64	57	48	62	60
	M	435,0	482,3	506,1	521,2	532,6
	SD	39,4	18,9	16,8	13,7	16,5
	r _{CC}	0,989	0,962	0,958	0,938	0,959
	SE _m (W)	16,2	13,2	11,9	11,6	11,1
	SE _m (SS)	12,1	10,1	9,1	9,0	8,5
Test 5: Fonologické zpracování	n	64	57	48	62	60
	M	462,3	487,4	507,2	512,5	518,1
	SD	15,1	19,1	12,7	12,8	12,8
	r _{CC}	0,977	0,987	0,980	0,981	0,982
	SE _m (W)	11,2	8,9	7,4	7,5	7,6
	SE _m (SS)	12,7	9,7	8,0	8,0	8,0
Test 6: Reprodukce příběhů	n	64	57	48	62	60
	M	489,2	498,9	505,4	506,1	509,4
	SD	12,0	10,9	8,8	12,2	7,6
	r _{CC}	0,969	0,971	0,957	0,973	0,952
	SE _m (W)	4,1	3,3	2,9	3,0	2,8
	SE _m (SS)	5,5	4,7	4,2	4,0	4,1
Test 7: Vizualizace	n	64	57	48	62	60
	M	486,9	495,5	502,7	505,2	508,4
	SD	11,9	10,3	10,1	11,3	9,1
	r _{CC}	0,957	0,970	0,943	0,974	0,930
	SE _m (W)	6,3	6,1	5,8	6,1	5,8
	SE _m (SS)	11,1	8,7	7,5	7,2	6,6

Test	Statistika	VĚK				
		5;0 – 6;11	7;0 – 8;11	9;0 – 10;11	11;0 – 12;11	13;0 – 14;11
Test 8: Všeobecné znalosti	n	64	57	48	62	60
	M	480,8	491,2	503,4	506,7	510,1
	SD	13,6	16,5	10,6	9,6	10,8
	r_{CC}	0,961	0,975	0,943	0,933	0,944
	SE_m (W)	7,2	6,7	6,3	6,2	6,5
	SE_m (SS)	7,6	7,0	6,7	6,6	6,9
Test 9: Formování konceptů	n	64	57	48	62	60
	M	471,4	491,4	502,4	506,9	514,8
	SD	27,9	23,0	16,4	26,9	28,8
	r_{CC}	0,990	0,989	0,982	0,992	0,992
	SE_m (W)	6,9	5,3	4,7	5,5	6,1
	SE_m (SS)	5,2	3,7	3,5	4,0	5,4
Test 10: Obrácené číselné řady	n	64	57	48	62	60
	M	445,5	487,7	510,1	516,1	528,2
	SD	40,1	27,9	18,1	25,0	17,1
	r_{CC}	0,992	0,988	0,979	0,989	0,978
	SE_m (W)	11,5	8,8	6,8	6,7	6,3
	SE_m (SS)	8,1	6,1	4,8	4,7	4,5
Rozšířená baterie						
Test 11: Hledání čísel	n	64	57	48	62	60
	M	420,1	487,4	524,9	536,6	549,9
	SD	50,1	36,2	22,0	28,7	15,6
	r_{CC}	0,992	0,987	0,975	0,986	0,958
	SE_m (W)	18,9	16,1	12,1	11,2	10,2
	SE_m (SS)	14,5	11,1	9,2	8,5	7,7
Test 12: Opakování pseudoslov	n	64	57	48	62	60
	M	483,9	491,2	499,2	503,8	510,1
	SD	19,7	17,6	13,6	18,0	13,4
	r_{CC}	0,986	0,983	0,971	0,983	0,969
	SE_m (W)	5,3	5,3	5,3	5,6	5,5
	SE_m (SS)	4,8	4,9	5,1	5,2	5,3
Test 13: Audio-vizuální učení	n	64	57	48	62	60
	M	482,6	493,0	503,8	505,2	510,3
	SD	15,2	14,5	9,9	11,5	7,3
	r_{CC}	0,977	0,984	0,979	0,982	0,977
	SE_m (W)	3,3	3,1	3,8	4,3	4,6
	SE_m (SS)	4,0	3,8	4,6	5,0	5,5
Test 14: Rozpoznávání obrázků	n	64	57	48	62	60
	M	487,6	492,3	500,6	501,1	506,3
	SD	12,5	14,3	9,4	15,1	11,1
	r_{CC}	0,962	0,971	0,937	0,974	0,952
	SE_m (W)	6,0	5,9	5,6	5,8	5,8
	SE_m (SS)	6,6	6,9	7,4	7,8	8,5

Test	Statistika	VĚK				
		5;0 – 6;11	7;0 – 8;11	9;0 – 10;11	11;0 – 12;11	13;0 – 14;11
Test 15: Analýza – Syntéza	n	64	57	48	62	60
	M	465,8	487,8	505,5	510,4	519,1
	SD	25,1	22,6	15,0	18,3	12,9
	r_{CC}	0,988	0,989	0,975	0,982	0,963
	SE_m (W)	6,8	5,5	5,6	5,9	6,1
	SE_m (SS)	6,3	4,7	4,8	5,1	5,4
Test 16: Řazení názvů a čísel	n	64	57	48	62	60
	M	468,4	486,9	506,7	510,6	519,3
	SD	18,3	19,0	14,9	20,8	19,2
	r_{CC}	0,974	0,980	0,969	0,983	0,981
	SE_m (W)	8,2	7,1	7,0	7,2	7,1
	SE_m (SS)	6,3	5,5	5,6	5,6	5,7
Test 17: Vyhledávání dvojic	n	64	57	48	62	60
	M	426,0	470,2	505,8	519,3	540,3
	SD	26,7	27,3	23,8	30,8	23,8
	r_{CC}	0,980	0,985	0,983	0,990	0,983
	SE_m (W)	13,9	11,2	9,8	9,6	9,3
	SE_m (SS)	9,5	7,7	6,8	5,9	6,1
Test 18: Paměť na slova	n	64	57	48	62	60
	M	476,3	488,0	497,2	504,6	515,1
	SD	20,6	20,7	16,9	24,0	20,8
	r_{CC}	0,980	0,980	0,972	0,986	0,981
	SE_m (W)	8,3	8,3	8,1	8,2	8,1
	SE_m (SS)	6,0	5,9	5,9	5,8	6,0

Pozn.: n – počet osob; M – průměr; SD – směrodatná odchylka; r_{CC} – odhad reliability; SE_m (W) – standardní chyba měření W-skóru; SE_m (SS) – standardní chyba měření skóru IQ

Test	Statistika	VĚK				
		15;0 – 16;11	17;0 – 18;11	19;0 – 20;11	21;0 – 25;11	26;0 – 30;11
Standardní baterie						
Test 1: Slovník	n	61	64	60	42	46
	M	522,5	531,2	538,2	532,3	537,2
	SD	11,2	17,2	14,0	14,3	17,2
	r _{CC}	0,956	0,979	0,968	0,971	0,978
	SE _m (W)	5,6	6,0	6,2	5,8	6,3
	SE _m (SS)	5,4	5,9	6,0	5,7	6,2
Test 2: Číselné řady	n	61	64	60	42	46
	M	532,3	536,3	542,0	538,7	544,9
	SD	17,4	17,2	12,3	20,7	21,0
	r _{CC}	0,981	0,981	0,964	0,986	0,986
	SE _m (W)	5,7	5,6	5,5	5,7	6,0
	SE _m (SS)	4,0	4,0	4,1	4,4	4,8
Test 3: Verbální pozornost	n	61	64	60	42	46
	M	519,0	523,9	529,0	523,6	527,3
	SD	11,5	17,6	12,5	13,3	13,8
	r _{CC}	0,959	0,982	0,965	0,969	0,971
	SE _m (W)	5,4	5,5	5,5	5,4	5,5
	SE _m (SS)	5,2	5,6	5,8	5,9	6,7
Test 4: Hledání písmen	n	61	64	60	42	46
	M	535,6	538,4	544,0	532,7	544,7
	SD	15,9	26,1	15,0	23,1	17,0
	r _{CC}	0,957	0,984	0,953	0,980	0,964
	SE _m (W)	10,8	10,8	10,4	10,9	10,4
	SE _m (SS)	8,4	8,2	8,0	8,3	8,1
Test 5: Fonologické zpracování	n	61	64	60	42	46
	M	522,7	527,4	532,7	529,5	533,1
	SD	12,6	14,8	9,6	15,7	13,1
	r _{CC}	0,982	0,986	0,967	0,987	0,980
	SE _m (W)	7,7	7,9	7,9	7,9	7,9
	SE _m (SS)	8,0	8,1	8,2	8,0	8,0
Test 6: Reprodukce příběhů	n	61	64	60	42	46
	M	511,6	512,5	514,2	509,6	514,1
	SD	8,2	9,3	7,8	9,4	8,6
	r _{CC}	0,959	0,964	0,949	0,963	0,965
	SE _m (W)	2,6	2,6	2,7	2,7	2,7
	SE _m (SS)	3,8	3,8	3,8	3,8	3,8
Test 7: Vizualizace	n	61	64	60	42	46
	M	510,8	510,7	513,9	512,5	518,2
	SD	10,3	11,8	9,1	10,9	15,0
	r _{CC}	0,943	0,955	0,926	0,948	0,967
	SE _m (W)	6,0	6,1	6,1	6,0	7,1
	SE _m (SS)	6,7	6,7	6,7	6,6	7,8

Test	Statistika	VĚK				
		15;0 – 16;11	17;0 – 18;11	19;0 – 20;11	21;0 – 25;11	26;0 – 30;11
Test 8: Všeobecné znalosti	n	61	64	60	42	46
	M	516,6	523,5	527,8	524,4	532,0
	SD	12,4	15,7	11,9	13,3	15,2
	r_{CC}	0,954	0,966	0,940	0,955	0,960
	SE_m (W)	6,8	8,0	8,1	7,6	8,8
	SE_m (SS)	7,2	8,5	8,6	8,1	9,3
Test 9: Formování konceptů	n	61	64	60	42	46
	M	517,3	518,4	522,9	521,3	523,3
	SD	22,0	23,0	14,5	20,6	29,0
	r_{CC}	0,988	0,988	0,970	0,982	0,990
	SE_m (W)	5,5	5,9	5,8	6,7	7,5
	SE_m (SS)	4,6	6,2	6,0	8,3	8,0
Test 10: Obrácené číselné řady	n	61	64	60	42	46
	M	525,4	534,0	539,0	538,8	545,5
	SD	17,6	21,8	20,6	21,3	23,8
	r_{CC}	0,979	0,987	0,986	0,986	0,989
	SE_m (W)	6,4	6,3	6,1	6,2	6,2
	SE_m (SS)	4,5	4,4	4,3	4,3	4,4
Rozšířená baterie						
Test 11: Hledání čísel	n	61	64	60	42	46
	M	554,1	556,3	560,6	552,2	565,7
	SD	13,6	23,5	11,4	22,5	11,4
	r_{CC}	0,946	0,981	0,927	0,980	0,930
	SE_m (W)	9,9	10,0	9,4	9,9	9,0
	SE_m (SS)	7,8	8,3	9,2	10,8	11,3
Test 12: Opakování pseudoslov	n	61	64	60	42	46
	M	512,2	511,4	516,0	517,2	518,2
	SD	13,7	15,4	14,9	19,5	15,8
	r_{CC}	0,971	0,977	0,974	0,985	0,977
	SE_m (W)	5,5	5,6	5,7	5,7	5,7
	SE_m (SS)	5,3	5,3	5,4	5,5	5,4
Test 13: Audio-vizuální učení	n	61	64	60	42	46
	M	507,6	509,6	511,3	508,1	512,0
	SD	9,0	11,3	6,3	12,2	11,5
	r_{CC}	0,975	0,983	0,969	0,984	0,986
	SE_m (W)	4,0	4,6	4,6	4,3	5,7
	SE_m (SS)	4,9	5,6	5,6	5,2	6,9
Test 14: Rozpoznávání obrázků	n	61	64	60	42	46
	M	506,9	507,1	510,3	509,1	508,9
	SD	10,7	10,6	10,0	11,1	13,6
	r_{CC}	0,948	0,948	0,940	0,951	0,966
	SE_m (W)	5,8	5,8	5,9	5,9	6,2
	SE_m (SS)	8,6	8,7	8,8	8,7	8,6

Test	Statistika	VĚK				
		15;0 – 16;11	17;0 – 18;11	19;0 – 20;11	21;0 – 25;11	26;0 – 30;11
Test 15: Analýza – Syntéza	n	61	64	60	42	46
	M	519,3	518,7	520,7	515,4	521,6
	SD	12,6	16,1	13,0	20,1	15,2
	r_{CC}	0,962	0,976	0,964	0,985	0,973
	SE_m (W)	6,0	6,1	6,1	6,0	6,1
	SE_m (SS)	5,3	5,4	5,4	5,4	5,7
Test 16: Řazení názvů a čísel	n	61	64	60	42	46
	M	524,1	532,1	537,6	533,4	539,9
	SD	16,7	21,3	19,9	20,4	20,2
	r_{CC}	0,974	0,984	0,982	0,983	0,982
	SE_m (W)	7,2	7,2	7,1	7,1	7,1
	SE_m (SS)	5,8	5,7	5,7	5,7	5,6
Test 17: Vyhledávání dvojic	n	61	64	60	42	46
	M	551,2	554,5	561,9	547,5	563,8
	SD	26,7	30,6	19,9	39,3	34,8
	r_{CC}	0,987	0,991	0,978	0,994	0,993
	SE_m (W)	8,9	8,7	8,5	8,9	8,3
	SE_m (SS)	5,5	5,1	4,8	4,1	4,3
Test 18: Paměť na slova	n	61	64	60	42	46
	M	519,1	526,5	530,6	529,1	534,4
	SD	18,6	21,0	19,6	22,8	25,4
	r_{CC}	0,976	0,981	0,978	0,984	0,986
	SE_m (W)	8,2	8,4	8,5	8,5	8,7
	SE_m (SS)	5,9	6,1	6,2	6,2	6,2

Pozn.: n – počet osob; M – průměr; SD – směrodatná odchylka; r_{CC} – odhad reliability; SE_m (W) – standardní chyba měření W-skóru; SE_m (SS) – standardní chyba měření skóru IQ

Test	Statistika	VĚK					
		31;0 – 40;11	41;0 – 50;11	51;0 – 60;11	61;0 – 70;11	71;0 – 80;11	medián
Standardní baterie							
Test 1: Slovník	n	93	90	60	42	19	0,971
	M	537,5	535,8	527,7	530,9	526,0	
	SD	17,1	16,7	18,1	18,8	12,1	
	r _{CC}	0,978	0,978	0,982	0,982	0,962	
	SE _m (W)	6,2	6,2	5,8	6,1	5,6	
	SE _m (SS)	6,1	6,0	5,7	5,9	5,4	
Test 2: Číselné řady	n	93	90	60	42	19	0,982
	M	542,4	536,9	537,0	535,2	526,8	
	SD	18,5	17,5	14,3	15,0	15,9	
	r _{CC}	0,983	0,982	0,973	0,976	0,977	
	SE _m (W)	5,8	5,6	5,5	5,5	5,7	
	SE _m (SS)	5,0	5,3	5,9	7,8	8,3	
Test 3: Verbální pozornost	n	93	90	60	42	19	0,971
	M	523,5	522,8	518,9	518,6	511,7	
	SD	18,6	14,4	16,2	18,1	12,4	
	r _{CC}	0,983	0,974	0,979	0,983	0,964	
	SE _m (W)	5,6	5,5	5,5	5,6	5,5	
	SE _m (SS)	7,2	7,1	6,4	6,2	4,7	
Test 4: Hledání písmen	n	93	90	60	42	19	0,966
	M	534,8	529,0	516,2	510,9	507,8	
	SD	20,9	18,1	18,3	20,0	21,1	
	r _{CC}	0,975	0,966	0,965	0,970	0,972	
	SE _m (W)	10,9	11,1	11,7	11,8	12,1	
	SE _m (SS)	8,4	8,6	8,9	9,1	9,0	
Test 5: Fonologické zpracování	n	93	90	60	42	19	0,982
	M	530,5	531,1	525,9	526,5	518,9	
	SD	17,5	14,7	16,2	14,0	14,0	
	r _{CC}	0,988	0,986	0,988	0,986	0,987	
	SE _m (W)	7,9	7,6	7,5	7,4	7,5	
	SE _m (SS)	7,5	7,0	6,7	6,3	6,2	
Test 6: Reprodukce příběhů	n	93	90	60	42	19	0,969
	M	510,2	510,5	503,0	505,6	503,8	
	SD	12,7	10,3	15,3	11,2	10,9	
	r _{CC}	0,975	0,972	0,980	0,971	0,973	
	SE _m (W)	2,8	2,7	3,0	2,8	2,8	
	SE _m (SS)	3,8	3,8	4,1	4,0	4,1	
Test 7: Vizualizace	n	93	90	60	42	19	0,955
	M	514,6	515,3	511,4	509,2	503,5	
	SD	13,0	11,4	10,8	9,0	9,5	
	r _{CC}	0,961	0,950	0,975	0,928	0,936	
	SE _m (W)	6,4	6,4	6,2	5,8	5,7	
	SE _m (SS)	7,1	7,0	6,6	7,0	6,9	

Test	Statistika	VĚK					
		31;0 – 40;11	41;0 – 50;11	51;0 – 60;11	61;0 – 70;11	71;0 – 80;11	medián
Test 8: Všeobecné znalosti	n	93	90	60	42	19	0,958
	M	533,5	534,9	527,4	534,3	533,4	
	SD	15,0	14,0	22,2	17,0	14,4	
	r _{CC}	0,958	0,951	0,981	0,965	0,955	
	SE _m (W)	9,0	9,2	8,6	9,5	9,1	
	SE _m (SS)	9,6	10,1	9,1	10,1	11,6	
Test 9: Formování konceptů	n	93	90	60	42	19	0,988
	M	519,6	516,1	513,1	511,0	491,9	
	SD	19,0	20,6	22,7	12,6	32,6	
	r _{CC}	0,981	0,985	0,988	0,970	0,994	
	SE _m (W)	6,1	5,8	5,4	4,7	5,6	
	SE _m (SS)	5,5	5,5	5,3	4,8	5,2	
Test 10: Obrácené číselné řady	n	93	90	60	42	19	0,986
	M	535,9	536,3	532,8	528,8	512,1	
	SD	27,4	18,1	26,2	18,4	21,0	
	r _{CC}	0,991	0,981	0,991	0,981	0,983	
	SE _m (W)	6,3	6,2	6,4	6,3	7,1	
	SE _m (SS)	4,4	4,4	4,4	4,4	5,1	
Rozšířená baterie							
Test 11: Hledání čísel	n	93	90	60	42	19	0,977
	M	555,2	556,0	543,4	540,7	533,6	
	SD	28,1	13,5	31,7	20,0	25,5	
	r _{CC}	0,988	0,946	0,989	0,973	0,982	
	SE _m (W)	9,6	9,8	10,6	10,9	11,6	
	SE _m (SS)	7,6	7,5	7,9	8,1	8,5	
Test 12: Opakování pseudoslov	n	93	90	60	42	19	0,983
	M	510,4	507,7	506,6	505,3	493,5	
	SD	21,6	16,0	19,6	12,3	18,9	
	r _{CC}	0,988	0,978	0,985	0,964	0,985	
	SE _m (W)	5,6	5,5	5,6	5,4	5,2	
	SE _m (SS)	5,3	5,3	5,3	5,2	5,0	
Test 13: Audio-vizuální učení	n	93	90	60	42	19	0,982
	M	507,1	505,0	503,6	500,8	490,8	
	SD	11,3	11,0	11,4	13,1	13,2	
	r _{CC}	0,985	0,984	0,984	0,982	0,982	
	SE _m (W)	4,5	4,3	4,0	3,7	2,9	
	SE _m (SS)	5,4	5,1	4,8	4,6	3,5	
Test 14: Rozpoznávání obrázků	n	93	90	60	42	19	0,952
	M	507,7	509,5	507,0	506,9	500,9	
	SD	14,5	9,1	15,6	8,1	10,6	
	r _{CC}	0,971	0,930	0,975	0,915	0,950	
	SE _m (W)	5,9	5,8	5,9	5,6	5,7	
	SE _m (SS)	7,8	7,9	9,6	9,8	9,2	

Test	Statistika	VĚK					
		31;0 – 40;11	41;0 – 50;11	51;0 – 60;11	61;0 – 70;11	71;0 – 80;11	medián
Test 15: Analýza – Syntéza	n	93	90	60	42	19	0,976
	M	515,9	514,1	511,2	509,7	500,4	
	SD	19,4	16,2	17,9	18,6	14,1	
	r_{CC}	0,983	0,977	0,981	0,982	0,973	
	SE_m (W)	6,1	5,9	5,9	5,9	5,4	
	SE_m (SS)	5,8	5,9	5,9	6,4	6,1	
Test 16: Řazení názvů a čísel	n	93	90	60	42	19	0,982
	M	531,5	533,4	525,8	526,5	516,1	
	SD	20,8	18,5	23,9	20,9	22,2	
	r_{CC}	0,983	0,979	0,987	0,984	0,986	
	SE_m (W)	7,3	7,2	7,3	7,2	7,0	
	SE_m (SS)	5,7	5,7	5,7	5,7	5,6	
Test 17: Vyhledávání dvojic	n	93	90	60	42	19	0,990
	M	548,8	548,7	511,8	519,7	504,7	
	SD	42,0	33,5	43,3	31,1	33,3	
	r_{CC}	0,995	0,992	0,994	0,990	0,991	
	SE_m (W)	8,8	8,8	10,2	9,6	10,3	
	SE_m (SS)	4,1	4,6	3,9	4,4	4,8	
Test 18: Paměť na slova	n	93	90	60	42	19	0,981
	M	523,5	520,6	516,7	515,3	507,8	
	SD	25,7	18,9	25,2	19,5	21,1	
	r_{CC}	0,987	0,977	0,987	0,979	0,982	
	SE_m (W)	8,5	8,2	8,3	8,1	8,1	
	SE_m (SS)	6,1	6,0	6,0	5,9	5,8	

Pozn.: n – počet osob; M – průměr; SD – směrodatná odchylka; r_{CC} – odhad reliability; SE_m (W) – standardní chyba měření W-skóru; SE_m (SS) – standardní chyba měření skóru IQ

Příloha C

Souhrnné popisné statistiky a údaje o reliabilitě pro klastrové skóry

Shluk	Statistika	VĚK				
		5;0 – 6;11	7;0 – 8;11	9;0 – 10;11	11;0 – 12;11	13;0 – 14;11
Obecná intelektová schopnost (GIA)	n	64	57	48	62	60
	M	462,6	489,3	505,5	512,3	518,3
	SD	18,6	13,3	9,8	9,4	9,8
	r_{CC}	0,971	0,955	0,927	0,922	0,930
	SE_m (W)	9,9	7,9	7,0	7,0	6,7
	SE_m (SS)	12,2	9,8	8,6	8,6	8,4
Zkrácená intelektová škála (BIA)	n	64	57	48	62	60
	M	457,0	488,4	506,9	514,7	520,6
	SD	24,2	17,4	12,2	11,8	11,5
	r_{CC}	0,984	0,978	0,961	0,958	0,958
	SE_m (W)	8,7	6,5	5,8	5,7	5,6
	SE_m (SS)	8,4	6,4	5,8	5,8	5,6
Kombinovaný skór Gf-Gc (GfGc)	n	64	57	48	62	60
	M	464,5	490,4	505,8	511,4	517,6
	SD	20,7	16,5	12,0	12,9	13,3
	r_{CC}	0,980	0,976	0,960	0,964	0,965
	SE_m (W)	8,1	6,4	5,8	5,9	6,1
	SE_m (SS)	8,3	6,5	5,9	6,1	6,3
Porozumění – Znalosti (Gc)	n	64	57	48	62	60
	M	477,8	492,7	505,9	510,1	515,0
	SD	15,8	14,7	11,2	9,9	10,7
	r_{CC}	0,968	0,969	0,952	0,940	0,946
	SE_m (W)	7,9	6,7	6,1	5,9	6,1
	SE_m (SS)	8,7	7,3	6,6	6,5	6,7
Fluidní usuzování (Gf)	n	64	57	48	62	60
	M	453,2	490,2	507,6	514,9	522,2
	SD	29,9	20,6	15,2	19,5	18,8
	r_{CC}	0,990	0,986	0,976	0,984	0,982
	SE_m (W)	8,3	6,0	5,4	5,9	6,0
	SE_m (SS)	6,1	4,5	4,2	4,5	4,7
Fluidní usuzování rozšířené (Gfe)	n	64	57	48	62	60
	M	457,4	489,4	506,9	513,4	521,2
	SD	25,8	19,6	12,9	18,1	15,5
	r_{CC}	0,987	0,984	0,967	0,982	0,974
	SE_m (W)	8,0	5,9	5,5	5,9	6,1
	SE_m (SS)	6,8	5,1	4,8	5,2	5,4
Krátkodobá pracovní paměť (Gwm)	n	64	57	48	62	60
	M	457,4	489,4	506,9	513,4	521,2
	SD	25,8	19,6	12,9	18,1	15,5
	r_{CC}	0,987	0,980	0,964	0,974	0,968
	SE_m (W)	8,0	5,9	5,5	5,9	6,1
	SE_m (SS)	6,8	5,1	4,8	5,2	5,4

Shluk	Statistika	VĚK				
		5;0 – 6;11	7;0 – 8;11	9;0 – 10;11	11;0 – 12;11	13;0 – 14;11
Krátkodobá pracovní paměť rozšířená (Gwm)	n	64	57	48	62	60
	M	460,4	487,0	506,8	512,9	521,3
	SD	23,9	18,1	12,7	16,3	13,9
	r _{CC}	0,983	0,976	0,960	0,975	0,967
	SE _m (W)	9,6	7,6	6,5	6,5	6,3
	SE _m (SS)	8,5	6,9	6,0	5,9	5,8
Rychlost kognitivního zpracování (Gs)	n	64	57	48	62	60
	M	431,5	477,3	506,9	521,2	537,4
	SD	31,0	21,3	18,6	19,6	17,8
	r _{CC}	0,984	0,973	0,969	0,972	0,967
	SE _m (W)	15,2	12,3	10,9	10,6	10,3
	SE _m (SS)	8,0	7,6	7,2	7,0	6,9
Auditivní zpracování (Ga)	n	64	57	48	62	60
	M	474,2	490,0	503,2	508,9	514,2
	SD	11,9	14,4	11,4	10,6	11,3
	r _{CC}	0,937	0,964	0,951	0,941	0,948
	SE _m (W)	8,8	7,4	6,4	6,6	6,6
	SE _m (SS)	10,3	8,6	7,5	7,6	7,7
Dlouhodobá paměť (Glr)	n	64	57	48	62	60
	M	485,3	495,3	503,3	504,4	509,9
	SD	10,7	10,6	8,7	12,7	8,8
	r _{CC}	0,965	0,973	0,964	0,978	0,960
	SE _m (W)	3,7	2,9	2,7	3,0	3,0
	SE _m (SS)	6,6	4,6	3,7	3,6	3,7
Vizuální zpracování (Gv)	n	64	57	48	62	60
	M	490,4	497,1	504,8	506,3	510,5
	SD	10,9	10,5	7,9	10,9	8,9
	r _{CC}	0,947	0,944	0,909	0,949	0,925
	SE _m (W)	6,2	6,0	5,7	6,0	5,8
	SE _m (SS)	9,3	8,8	8,6	8,8	8,8
Kvantitativní usuzování (RQ)	n	64	57	48	62	60
	M	449,9	487,8	508,6	516,0	523,8
	SD	28,8	21,5	13,3	16,9	13,0
	r _{CC}	0,989	0,987	0,967	0,979	0,965
	SE _m (W)	8,3	6,1	5,8	6,0	5,9
	SE _m (SS)	6,4	5,0	4,9	5,2	5,3
Číselné dovednosti (N)	n	64	57	48	62	60
	M	429,2	483,0	512,9	521,8	534,5
	SD	38,7	26,2	17,2	24,3	14,0
	r _{CC}	0,989	0,980	0,967	0,984	0,956
	SE _m (W)	15,8	13,2	9,8	9,3	8,5
	SE _m (SS)	12,9	10,6	8,0	7,5	6,9

Shluk	Statistika	VĚK				
		5;0 – 6;11	7;0 – 8;11	9;0 – 10;11	11;0 – 12;11	13;0 – 14;11
Rychlost vnímání (P)	n	64	57	48	62	60
	M	420,5	477,8	508,4	521,8	534,2
	SD	43,0	26,2	18,7	18,6	15,1
	r_{CC}	0,990	0,978	0,966	0,967	0,953
	SE_m (W)	17,7	14,9	12,0	11,4	10,7
	SE_m (SS)	15,8	15,3	11,7	10,1	8,9
Kognitivní efektivita (CE)	n	64	57	48	62	60
	M	441,4	485,2	508,1	518,8	530,5
	SD	33,4	18,7	14,6	15,8	13,3
	r_{CC}	0,987	0,966	0,955	0,962	0,949
	SE_m (W)	14,2	11,4	9,7	9,5	9,0
	SE_m (SS)	2,0	9,7	8,4	8,2	7,7
Kognitivní efektivita rozšířená (CEe)	n	64	57	48	62	60
	M	438,4	482,3	507,6	517,9	528,2
	SD	31,8	19,6	14,0	14,6	12,3
	r_{CC}	0,985	0,968	0,951	0,957	0,943
	SE_m (W)	14,6	12,0	9,6	9,2	8,6
	SE_m (SS)	15,4	13,6	10,6	9,4	8,6

Pozn.: n – počet osob; M – průměr; SD – směrodatná odchylka; r_{CC} – odhad reliability; SE_m (W) – standardní chyba měření W-skóru; SE_m (SS) – standardní chyba měření skóru IQ

Shluk	Statistika	VĚK				
		15;0 – 16;11	17;0 – 18;11	19;0 – 20;11	21;0 – 25;11	26;0 – 30;11
Obecná intelektová schopnost (GIA)	n	61	64	60	42	46
	M	521,5	525,4	530,2	525,1	531,1
	SD	9,7	13,0	8,7	12,5	12,8
	r_{CC}	0,929	0,960	0,911	0,957	0,957
	SE_m (W)	6,7	6,8	6,7	6,8	7,0
	SE_m (SS)	8,3	8,6	8,6	8,6	9,2
Zkrácená intelektová škála (BIA)	n	61	64	60	42	46
	M	524,1	529,9	535,9	531,0	535,9
	SD	11,1	15,0	11,2	13,7	15,5
	r_{CC}	0,955	0,975	0,954	0,970	0,975
	SE_m (W)	5,6	5,7	5,8	5,7	6,0
	SE_m (SS)	5,7	5,9	6,0	6,0	6,6
Kombinovaný skóre Gf-Gc (GfGc)	n	61	64	60	42	46
	M	522,1	527,2	532,6	529,0	534,2
	SD	12,3	14,3	9,7	15,1	17,6
	r_{CC}	0,960	0,967	0,928	0,970	0,975
	SE_m (W)	6,0	6,6	6,6	6,6	7,4
	SE_m (SS)	6,2	6,9	6,9	6,9	8,0
Porozumění – Znalosti (Gc)	n	61	64	60	42	46
	M	520,7	528,4	534,1	529,5	535,7
	SD	11,1	15,8	12,2	13,0	15,2
	r_{CC}	0,949	0,971	0,950	0,959	0,965
	SE_m (W)	6,3	7,1	7,3	6,8	7,7
	SE_m (SS)	6,9	7,9	8,0	7,5	8,5
Fluidní usuzování (Gf)	n	61	64	60	42	46
	M	525,4	528,0	533,1	530,6	534,7
	SD	16,5	16,6	11,1	19,5	22,9
	r_{CC}	0,979	0,978	0,951	0,983	0,986
	SE_m (W)	5,6	5,8	5,8	6,4	7,0
	SE_m (SS)	4,4	4,6	4,6	5,3	6,1
Fluidní usuzování rozšířené (Gfe)	n	61	64	60	42	46
	M	523,4	524,9	529,0	525,6	530,4
	SD	13,7	15,0	10,3	18,3	19,4
	r_{CC}	0,969	0,973	0,943	0,981	0,982
	SE_m (W)	5,8	6,0	5,9	6,3	6,7
	SE_m (SS)	5,1	5,3	5,3	5,8	6,6
Krátkodobá pracovní paměť (Gwm)	n	61	64	60	42	46
	M	523,4	524,9	529,0	525,6	530,4
	SD	13,7	15,0	10,3	18,3	19,4
	r_{CC}	0,967	0,981	0,975	0,976	0,981
	SE_m (W)	5,8	6,0	5,9	6,3	6,7
	SE_m (SS)	5,1	5,3	5,3	5,8	6,6

Shluk	Statistika	VĚK				
		15;0 – 16;11	17;0 – 18;11	19;0 – 20;11	21;0 – 25;11	26;0 – 30;11
Krátkodobá pracovní paměť rozšířená (Gwm)	n	61	64	60	42	46
	M	522,5	529,7	534,9	531,6	537,3
	SD	13,3	18,1	15,9	15,9	17,3
	r _{CC}	0,964	0,981	0,975	0,975	0,979
	SE _m (W)	6,4	6,4	6,3	6,3	6,3
	SE _m (SS)	5,9	5,8	5,8	5,8	5,8
Rychlost kognitivního zpracování (Gs)	n	61	64	60	42	46
	M	544,3	547,4	554,0	541,1	555,2
	SD	19,8	26,7	16,2	30,0	21,1
	r _{CC}	0,975	0,986	0,963	0,989	0,979
	SE _m (W)	9,9	9,8	9,5	10,0	9,4
	SE _m (SS)	6,6	6,4	6,4	6,4	6,3
Auditivní zpracování (Ga)	n	61	64	60	42	46
	M	517,8	520,1	525,1	523,8	526,8
	SD	11,2	12,8	10,5	15,7	13,7
	r _{CC}	0,946	0,958	0,937	0,972	0,963
	SE _m (W)	6,7	6,8	6,9	6,9	6,9
	SE _m (SS)	7,8	7,9	8,1	8,1	8,1
Dlouhodobá paměť (Glr)	n	61	64	60	42	46
	M	509,5	510,3	512,0	508,0	512,3
	SD	8,4	9,6	7,2	10,1	12,0
	r _{CC}	0,962	0,969	0,946	0,973	0,976
	SE _m (W)	2,7	2,8	2,8	2,7	3,2
	SE _m (SS)	3,2	3,3	3,3	3,2	3,9
Vizuální zpracování (Gv)	n	61	64	60	42	46
	M	512,0	512,0	515,3	513,9	516,7
	SD	9,1	9,3	8,0	9,4	13,3
	r _{CC}	0,928	0,930	0,905	0,931	0,961
	SE _m (W)	5,9	6,0	6,0	6,0	6,7
	SE _m (SS)	9,0	9,0	9,1	9,1	10,1
Kvantitativní usuzování (RQ)	n	61	64	60	42	46
	M	525,9	527,6	531,4	527,1	533,3
	SD	12,6	14,7	10,6	18,6	16,8
	r _{CC}	0,963	0,973	0,948	0,983	0,978
	SE _m (W)	5,9	5,9	5,8	5,9	6,1
	SE _m (SS)	5,2	5,4	5,4	5,7	6,3
Číselné dovednosti (N)	n	61	64	60	42	46
	M	535,1	540,6	545,2	541,3	551,0
	SD	13,4	19,3	13,1	18,6	16,4
	r _{CC}	0,954	0,977	0,953	0,976	0,971
	SE _m (W)	8,3	8,4	7,9	8,3	7,7
	SE _m (SS)	6,8	6,6	6,5	6,8	6,3

Shluk	Statistika	VĚK				
		15;0 – 16;11	17;0 – 18;11	19;0 – 20;11	21;0 – 25;11	26;0 – 30;11
Rychlost vnímání (P)	n	61	64	60	42	46
	M	537,7	540,3	545,2	535,7	548,1
	SD	14,0	24,2	12,1	22,5	13,4
	r_{CC}	0,946	0,982	0,932	0,979	0,945
	SE_m (W)	10,4	10,4	9,9	10,4	9,7
	SE_m (SS)	8,0	7,9	7,7	7,8	9,0
Kognitivní efektivita (CE)	n	61	64	60	42	46
	M	530,6	536,3	541,6	535,9	545,2
	SD	14,6	20,1	14,5	18,6	18,2
	r_{CC}	0,958	0,978	0,959	0,975	0,974
	SE_m (W)	8,9	8,8	8,6	8,8	8,6
	SE_m (SS)	7,7	7,6	7,4	7,6	7,4
Kognitivní efektivita rozšířená (CEe)	n	61	64	60	42	46
	M	529,7	534,3	539,3	533,2	542,0
	SD	12,0	18,5	11,3	16,3	14,2
	r_{CC}	0,941	0,975	0,936	0,968	0,960
	SE_m (W)	8,5	8,5	8,1	8,4	8,0
	SE_m (SS)	7,7	8,0	7,9	7,8	8,7

Pozn.: n – počet osob; M – průměr; SD – směrodatná odchylka; r_{CC} – odhad reliability; SE_m (W) – standardní chyba měření W-skóru; SE_m (SS) – standardní chyba měření skóru IQ

Shluk	Statistika	VĚK					Medián
		31;0 – 40;11	41;0 – 50;11	51;0 – 60;11	61;0 – 70;11	71;0 – 80;11	
Obecná intelektová schopnost (GIA)	n	93	90	60	42	19	0,946
	M	527,4	525,6	519,7	519,2	513,6	
	SD	12,8	11,5	11,4	11,7	9,5	
	r_{CC}	0,957	0,948	0,946	0,949	0,923	
	SE_m (W)	7,0	6,9	7,0	6,9	7,0	
	SE_m (SS)	9,3	9,4	9,6	10,5	10,0	
Zkrácená intelektová škála (BIA)	n	93	90	60	42	19	0,963
	M	533,9	531,3	527,3	527,7	521,0	
	SD	14,0	12,9	12,4	13,6	9,8	
	r_{CC}	0,970	0,965	0,963	0,969	0,942	
	SE_m (W)	5,9	5,8	5,6	5,7	5,6	
	SE_m (SS)	6,7	6,8	7,0	8,0	7,5	
Kombinovaný skór Gf-Gc (GfGc)	n	93	90	60	42	19	0,965
	M	533,1	530,8	526,2	527,7	519,4	
	SD	14,8	14,2	15,1	12,3	12,4	
	r_{CC}	0,967	0,965	0,970	0,954	0,954	
	SE_m (W)	7,1	7,0	6,6	6,8	6,9	
	SE_m (SS)	7,7	7,7	7,4	8,2	7,8	
Porozumění – Znalosti (Gc)	n	93	90	60	42	19	0,959
	M	536,6	536,5	528,7	533,7	530,8	
	SD	15,0	14,2	17,7	17,0	10,8	
	r_{CC}	0,964	0,960	0,975	0,971	0,933	
	SE_m (W)	7,8	7,9	7,4	8,1	7,6	
	SE_m (SS)	8,6	8,7	8,1	8,9	8,4	
Fluidní usuzování (Gf)	n	93	90	60	42	19	0,979
	M	531,6	527,1	525,7	523,7	510,0	
	SD	17,6	17,4	15,2	11,9	20,1	
	r_{CC}	0,979	0,980	0,975	0,964	0,985	
	SE_m (W)	6,1	5,8	5,6	5,1	5,8	
	SE_m (SS)	5,7	5,7	5,7	6,9	7,0	
Fluidní usuzování rozšířené (Gfe)	n	93	90	60	42	19	0,976
	M	526,4	522,8	520,9	519,1	506,8	
	SD	16,3	15,8	15,4	11,8	16,8	
	r_{CC}	0,976	0,976	0,976	0,961	0,979	
	SE_m (W)	6,1	5,9	5,7	5,4	5,7	
	SE_m (SS)	6,3	6,1	6,0	6,9	6,9	
Krátkodobá pracovní paměť (Gwm)	n	93	90	60	42	19	0,975
	M	526,4	522,8	520,9	519,1	506,8	
	SD	16,3	15,8	15,4	11,8	16,8	
	r_{CC}	0,985	0,970	0,978	0,978	0,972	
	SE_m (W)	6,1	5,9	5,7	5,4	5,7	
	SE_m (SS)	6,3	6,1	6,0	6,9	6,9	

Shluk	Statistika	VĚK					Medián
		31;0 – 40;11	41;0 – 50;11	51;0 – 60;11	61;0 – 70;11	71;0 – 80;11	
Krátkodobá pracovní paměť rozšířená (Gwm)	n	93	90	60	42	19	0,975
	M	530,0	530,5	525,6	524,3	513,0	
	SD	19,2	14,4	18,2	17,0	16,7	
	r_{CC}	0,982	0,970	0,980	0,978	0,976	
	SE_m (W)	6,4	6,3	6,4	6,4	6,7	
	SE_m (SS)	5,8	5,8	5,8	5,9	6,1	
Rychlost kognitivního zpracování (Gs)	n	93	90	60	42	19	0,979
	M	542,8	539,9	515,0	516,3	507,2	
	SD	28,2	23,5	27,1	23,9	25,6	
	r_{CC}	0,987	0,982	0,985	0,981	0,983	
	SE_m (W)	9,9	10,0	11,0	10,8	11,3	
	SE_m (SS)	6,5	6,6	7,2	7,1	7,4	
Auditivní zpracování (Ga)	n	93	90	60	42	19	0,958
	M	522,1	520,4	517,6	516,7	506,5	
	SD	16,0	13,7	14,2	12,4	14,4	
	r_{CC}	0,973	0,964	0,966	0,958	0,969	
	SE_m (W)	6,9	6,7	6,7	6,5	6,4	
	SE_m (SS)	7,9	7,8	7,8	7,5	7,5	
Dlouhodobá paměť (Glr)	n	93	90	60	42	19	0,973
	M	507,4	506,4	501,3	501,3	497,2	
	SD	11,3	10,7	14,3	10,6	10,1	
	r_{CC}	0,976	0,976	0,984	0,977	0,974	
	SE_m (W)	2,9	2,7	2,9	2,6	2,6	
	SE_m (SS)	3,6	3,5	3,6	3,7	3,6	
Vizuální zpracování (Gv)	n	93	90	60	42	19	0,931
	M	514,3	515,5	512,4	511,2	505,3	
	SD	11,1	8,8	11,4	7,0	7,7	
	r_{CC}	0,949	0,920	0,952	0,884	0,903	
	SE_m (W)	6,2	6,1	6,0	5,7	5,7	
	SE_m (SS)	9,3	9,3	8,9	8,6	8,6	
Kvantitativní usuzování (RQ)	n	93	90	60	42	19	0,970
	M	529,2	525,5	524,2	522,5	513,7	
	SD	16,1	15,5	13,8	13,0	12,9	
	r_{CC}	0,977	0,976	0,970	0,966	0,967	
	SE_m (W)	6,0	5,8	5,7	5,7	5,6	
	SE_m (SS)	6,7	6,5	6,3	6,7	5,6	
Číselné dovednosti (N)	n	93	90	60	42	19	0,971
	M	541,0	541,6	533,6	530,2	518,3	
	SD	24,7	12,5	26,0	16,7	20,7	
	r_{CC}	0,986	0,948	0,987	0,968	0,977	
	SE_m (W)	8,1	8,2	8,8	8,9	9,7	
	SE_m (SS)	6,5	6,7	7,1	7,3	7,7	

Shluk	Statistika	VĚK					
		31;0 – 40;11	41;0 – 50;11	51;0 – 60;11	61;0 – 70;11	71;0 – 80;11	Medián
Rychlost vnímání (P)	n	93	90	60	42	19	0,967
	M	537,9	535,5	522,7	518,7	513,6	
	SD	20,5	15,0	20,9	19,4	22,1	
	r_{CC}	0,975	0,953	0,974	0,970	0,975	
	SE_m (W)	10,3	10,5	11,2	11,4	11,9	
	SE_m (SS)	9,2	8,7	8,3	10,5	10,1	
Kognitivní efektivita (CE)	n	93	90	60	42	19	0,964
	M	535,5	532,8	524,6	520,0	510,0	
	SD	19,3	14,1	16,2	16,7	18,8	
	r_{CC}	0,976	0,955	0,964	0,966	0,971	
	SE_m (W)	8,9	9,0	9,4	9,5	9,9	
	SE_m (SS)	7,6	7,8	8,1	8,2	8,2	
Kognitivní efektivita rozšířená (CEe)	n	93	90	60	42	19	0,960
	M	533,5	532,2	524,0	520,9	512,5	
	SD	18,6	11,9	16,4	15,9	17,2	
	r_{CC}	0,975	0,940	0,966	0,964	0,967	
	SE_m (W)	8,4	8,5	9,0	9,1	9,6	
	SE_m (SS)	8,8	8,5	8,3	9,7	9,0	

Pozn.: n – počet osob; M – průměr; SD – směrodatná odchylka; r_{CC} – odhad reliability; SE_m (W) – standardní chyba měření W-skóru; SE_m (SS) – standardní chyba měření skóru IQ

