

## POSUDEK NA DIZERTAČNÍ PRÁCI

**Název práce: Neurální koreláty aritmetických funkcí**

**Autorka práce: Mgr. Michala Plassová**

**Vedoucí práce: Mgr. Michal Vavrečka, Ph.D.**

**Autor posudku: Mgr. Hynek Cígler, Ph.D.**

Předložená dizertační práce se zabývá neurálními koreláty aritmetických a prenumerických funkcí, zejména pak aproximálním numerickým systémem (ANS). Hned úvodem bych chtěl uvést, že se mi tento posudek v žádném případě nepíše snadno – mj. proto, že jsem seznámen s autorčinou publikací v časopise *Pedagogika* (Plassová, 2017), která se mi velice líbila, a těšil jsem se, až si budu moci přečíst celou dizertaci. Velmi oceňuji práci, kterou autorka odvedla a která je po mnoha stránkách kvalitní, bohužel však předložený text a vlastně i metodologie výzkumu obsahuje zásadní nedostatky. I přes značné sympatie k tématu, nasazení i výzkumnému záměru (velmi obdivuji např. nahrání scének s kolegy z amatérského divadla) proto nemohu napsat příznivý posudek. Zároveň chci jednoznačně deklarovat, že se necítím být odborníkem ani na neuropsychologii a neuroanatomii, ani na zpracování EEG dat, a proto pro mě bylo porozumění některým pasážím poměrně náročné (ke kvalitě nastavení EEG přístroje a základním datovým úpravám se nevyjadřuji vůbec, líbí se mi, že se s tím autorka vypořádala). Alespoň stran statistického zpracování jsem nicméně své postřehy (anonymně) konzultoval se zkušenějším kolegou.

Samotný teoretický úvod předkládá poměrně rozsáhlý přehled neurálních korelátů matematických a prematematických funkcí včetně jejich neuroanatomie a lokalizace. Zvýšenou pozornost pak věnuje právě ANS, zejména u dětí, a představuje i základy vyšetření prostřednictvím EEG. Obě oblasti jsou propojeny s využitím předchozích empirických výsledků. Vzhledem k mým mezerám v neuropsychologii nejsem schopný posoudit úplnost a adekvátnost předložených informací, které mi však (možná právě proto) připadaly velmi zajímavé a myslím, že šlo o informačně kvalitní přehled.

Bohužel i přes svou obsáhlost má tento úvod určité nedostatky. Prvním z nich je místy hůře navazující text, což je nepříliš přívětivé pro čtenáře; některé informace se částečně duplikují v různých kapitolách, což autorka sama občas reflektuje. Dvě několikastránkové pasáže přejeté z vlastní studie (Plassová, 2017) dále rozbíjejí přirozený tok textu a působí lehce nepatřičně (výtka není směřovaná na samotné převzetí textu, to považuji za přirozenou součást dizertačního projektu). Za poměrně nevhodnou pak považuji volbu českého překladu vlastní zkoušky ANS jako „aproximální numerický trénink“ v případě, že jde vlastně jen o samotné měření dané schopnosti, test či praktickou diskriminaci množstevních podnětů ve smyslu samotné aktivity (bez tréninkového cíle jako takového). Díky tomu je např. výzkumná otázka při prvním přečtení poměrně nejednoznačná, celkově jde ale spíše jen o detail.

Masarykova univerzita, Fakulta sociálních studií

Joštova 218/10, 602 00 Brno, Česká republika

T: +420 549 49 1910, E: info@fss.muni.cz, www.fss.muni.cz

Bankovní spojení: KB Brno-město, ČÚ: 85636621/0100, IČ: 00216224, DIČ: CZ00216224

Zásadní nedostatky však spatřuji zejména ve vlastním empirickém výzkumu. Za prvé výzkumná otázka nenavazuje zřetelně na předchozí teoretický úvod, výzkum není propojen s teoretickou rešerší. Hypotézy nejsou zdůvodněny a působí ad hoc dojmem. Ještě zásadnější je pak fakt, že byly evidentně formulovány až po zahájení analýz, kdy bylo jasné, že ve vzorku nejsou děti s podprůměrným výsledkem v Stanford-Binetově (SB) testu dle jeho norem (příčemž by navíc stačilo děti rozdělit dle mediánu v daném vzorku a normy vůbec nepoužívat). Stejně tak rozdělení hypotéz podle obtížnosti daných úloh nemá žádnou oporu v teorii a je pouze produktem prvotní analýzy dat, což však není jakkoliv reflektováno.

Použití samotného SB testu je rovněž vysoce problematické, zvláště pokud jsou interpretovány zastaralé a navíc (zřejmě) zahraniční normy, a to bez jakékoliv reflexe. Bylo přitom možné použít řadu jiných, recentních a v ČR pro danou věkovou skupinu adaptovaných testů (např. IDS, SON-R), nebo alespoň nevyužívat naprosto nevhodné normy a pracovat s hrubými skóry. Nelze souhlasit, že „*Stanford-Binetův inteligenční test [je] vhodný pro předškoláky*“ (s. 72), přinejmenším v kombinaci zvolené verze, norem a aktuálního českého prostředí. Chybí mi rovněž deskriptivy přímo na úrovni IQ nebo hrubého skóru a nikoliv arbitrárních pásem, hlavně však považuji za problematické či dokonce neetické sdělení inteligenčního pásma naměřeného tímto testem rodičům právě vzhledem k zásadní nevhodnosti použitých norem (nehledě na jejich zastaralost).

Zásadní výhrady mám pak k designu vlastního experimentu. Není mi jasný důvod, proč byly **vyřazeny chybné odpovědi z analýz** (a to navíc v rozporu s předchozími studiemi, jak sama autorka přiznává na s. 125); navíc není jisté, o jak velký podíl dat vlastně šlo. Autorka rovněž používá pro definici obtížnosti úloh „log“, aniž by jej vysvětlila – nehledě na to, že samotný překlad je nevhodný, a vhodnější by bylo použít buď běžný termín „logit“, nebo „logaritmus rozdílu“ ve shodě s „log-difference“ Parka a Brannonové (2014). První setkání s jednotkou je na s. 69: „... v původním experimentu Parka a Brannonové (2014) byly obtížnosti na dospělých lidech pro děti příliš obtížné (log 1.5 až 1.1).“ Autorka jednotku vysvětluje až na s. 75: „Kupříkladu při logaritmu 1.1 se množiny liší pouze o dva prvky, tj. je kupříkladu 51 oranžových rybek a 49 modrých.“ Na s. 119 pak v tab. 9 uvádí přepočtení poměru, kdy rozdíl 26 ryb ze 100 odpovídá logu 2,3, zatímco rozdíl dvou ryb logu 1,1. Jsem přesvědčen, že **přepočtení poměru na logy je chybné, a i nejjednodušší úlohy použité autorkou jsou výrazně náročnější než ty nejobtížnější** (s log = 1,1) použité ve studii Parka a Brannonové (2014) na dospělých osobách. Pokud mám pravdu, je zcela evidentní, že děti v řešení úloh selhávaly – byly jim totiž administrovány úlohy naprosto nepřiměřeně obtížné, prakticky neřešitelné i dospělými osobami. Je tedy otázkou, zda se ANS vlastně mohl při takto nepatrných poměrech projevit a být zaznamenatelný na EEG. V kombinaci s vyřazením neúspěšných úloh tak jde o zásadní metodologický nedostatek.

Další nedostatky pak spatřuji ve statistické analýze. Autorka nepopisuje blíže postup analýzy dat, resp. srovnání obou skupin z hlediska jejich EEG výsledků. Není tedy jisté, zda jsou srovnávány průměrné hodnoty jednotlivých trialů napříč skupinami, nebo zda jsou srovnávány přímo tyto trialy. Druhá metoda by prudce nadhodnotila počet signifikantních výsledků, avšak bez možnosti interpretací ohledně rozdílů skupin.

Zejména však autorka sama uvádí, že **p-hodnoty jsou bez korekce** (např. s. 86). Podíváme-li se však na veškeré grafy (většina obr. 31–61), vyznačené oblasti se signifikantním rozdílem zabírají většinou přibližně (odhadem) 5 % sledovaného časového rozmezí, což plně odpovídá náhodným výsledkům při zvolené hladině významnosti  $\alpha = 0,05$ . Vzhledem k určité volnosti při umístění jednotlivých „vln“ do konkrétního časového intervalu je tedy vysoce pravděpodobné, že **přinejmenším většina výsledků jsou falešná pozitiva**.

Autorka mezi hypotézy zařadila rovněž vztah mezi obtížností úlohy a reakčním časem, což je pochopitelné, ale i zde chybí vazba na teorii (nehledě na to, že převod původního „logu“ na rozdíl nemá vliv na velikost korelace, protože byla použita lineární transformace, viz tab. 9 na s. 119). Nicméně samotný výsledek je velice překvapivý: **náročnější úlohy byly řešeny v kratším čase**, což odporuje teoretickým předpokladům zmíněným samotnou autorkou – viz např. s. 34: „... (například reakční čas je kratší a přesnost vyšší, když posuzují, zda je více 12 modrých teček oproti 3 žlutým, než když posuzují, zda je více 6 modrých teček oproti 3 žlutým) (Haist et al., 2015),“ a dále. Autorka se však k tomuto rozporu oproti teoretickým předpokladům téměř nevyjadřuje, a sílu a směr vztahu reakčního času a náročnosti úloh nikde neinterpretuje (výjimkou je s. 128, viz níže). Dle mého názoru je pozorovaný vztah prostým důsledkem vyřazení neúspěšných pokusů.

Závěrem se pak domnívám, že u některých statistických testů byl volen chybný postup, který může vést k určitému zkreslení při interpretaci. Příkladem může být srovnání reakčního času obou skupin na s. 119, kdy byl použit párový t-test pro 13 průměrných časů získaných v obou skupinách separátně (navíc není jisté, jak a zda byl tento průměr vážen podle počtu trialů jednotlivých dětí, což může značně ovlivnit výsledné průměry). Domnívám se, že tento t-test není rozdílem populačních průměrů obou skupin (tedy s vyšším IQ,  $n = 8$ , a nižším IQ,  $n = 9$ ), ale testem jejich výběrových průměrů.

Dalšími drobnějšími nedostatky výsledkové části jsou:

- Outputy exportované s největší pravděpodobností z programu Jamovi jsou označeny jako výstup programu Statistica (s. 119).
- Přesné p-hodnoty jsou reportovány chybně s pomocí nerovnítko (např. „ $p < 0.014$ “, s. 119); nerovnítko se používá při srovnání p-hodnoty s arbitrárně zvolenou hladinou spolehlivosti (srov. např. APA, 2010, s. 114).
- Autorka bez vysvětlení variuje s hladinou signifikance alfa (různé kritické hodnoty na různých místech práce), což může nicméně být důsledek předchozího bodu.
- Neobratné reportování statistik ve výsledcích i teorii, např. „*Významným přínosem je práce Templové a Posnera (1998) na dvaatřiceti ( $n=13$ ) pětiletých dětech (mean 5.4 let).*“ (s. 55; v tomto konkrétním případě navíc nesedí slovní a číselný počet.)

Rovněž i diskuze obsahuje dílčí nedostatky nezmiňené dříve. Např. nelze jednoznačně souhlasit s nízkou variabilitou dětí z hlediska SB testu (srov. s. 124) – rozložení popsané v tabulce 4 (s. 70) nemusí být až tolik odlišné od očekávané populační variability. Zásadnějším nedostatkem je pak interpretace výsledků, které nebyly součástí hypotéz (např. výsledky na elektrodě N1, viz s. 125). Interpretace dodatečných zjištění není sice problematická sama o sobě, tato explorační zjištění je však nezbytné odlišit od konfirmačních testů hypotéz.

Dále pak závěr „*V rámci získaného reakčního času nacházíme již uvedenou výraznou korelaci mezi délkou RT a obtížností úlohy u obou skupin, kdy rozdíl mezi skupinami (rozdíl 160 ms, Cohen  $d = -1.03$ ,  $p < 0.003$ ) přispívá k již nastiněnému vztahu rychlosti nesymbolického numerického zpracování s výkonem v inteligenčním testu.*“ (s. 128) nedává smysl. Autorka na jednu stranu deklaruje, že „chytřejší“ děti měly nižší průměrný reakční čas, zároveň však obtížnější úlohy byly řešeny v kratším čase ( $r = .87$ , viz s. 119). **Tyto závěry jsou v kontradikci, což ovšem není reflektováno**, a autorčina interpretace tak nedává věcně smysl.

V celé práci pak shledávám další nedostatky (seznam není netříděný a řazený):

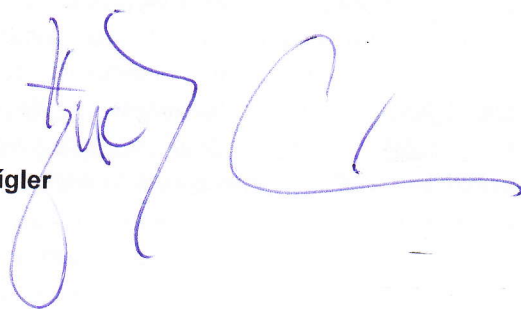
- Autorka přebrala celou řadu ilustrací, z nichž určitá část nebyla ve volné licenci. Autorka neuvádí, zda získala povolení k jejich převzetí a reprintu, což může být problematické z hlediska copyrightu.
- Autorka přechyluje jména výzkumnic, přechylování však není sjednocené (srov. popis obr. 2 na s. 24, „In Cantlon, 2006“ s referencí 21, s. 132: „Cantlonová, J. F., Brannonová, E. M., Carter, E. J., & Pelphrey, K. A. (2006). *Functional imaging of numerical processing in adults and 4-y-old children. PLoS Biology*, 4 (5), e125.“ Rovněž pak převzatý **obr. 2 se v citované studii Cantlonové nevyskytuje** a není tedy zřejmé, odkud byl převzat.
- Autorka zaměňuje velikost efektu a jeho signifikance, viz např.: „Autoři zjistili výraznou korelaci ( $p < 0.005$ )...“ (s. 27), na mnoha místech jsou uváděny p-hodnoty předchozích zjištění, namísto mnohem užitečnějších velikostí efektů (např. s. 35 dole).
- Považuji za vysoce nevhodné a eticky problematické, že je jmenovitě uvedená škola, která odmítla spolupráci (viz s. 69), a nebyla zachována její anonymita.
- Autorka dvakrát zmiňuje termín **prekognice** (s. 81 a s. 92). Na s. 81 se navíc odkazuje na publikaci Lucka (2014) *An Introduction to the Event-Related Potential Technique*. V ní se mi však termín dohledat nepodařilo. Není mi jasné, co autorka termínem myslela.
- Rozdíly v číslování tabulek v textu a v názvech tabulek.
- Liší se počty použitých trialů u nadprůměrných a průměrných dětí uvedených v textu (s. 86, 96 a 107) při srovnání s tab. 7.

Závěrem bych přesto rád pochválil dílčí části práce. Vyšetření malých dětí s EEG muselo být velmi náročné, což dokládá i nízký počet EEG studií na dětech v dané oblasti. Autorka si rovněž dala práci s konstrukcí podnětového materiálu, odhlédneme-li od chybně zvolených poměrů, např. kontrola hustoty s použitím Fibonacciho mřížky je jistě dobrým krokem. Stejně tak nastudování a rešerše neuropsychologické literatury musela být vysoce náročná.

I přes tato pozitiva však musím konstatovat, že předložená diplomová práce **neodpovídá úrovni běžně požadované na doktorské studium**. Zásadní nedostatky jsou v metodologii a vyhodnocení výzkumných dat. Dle mého názoru byl podnětový materiál zvolen chybně, což vedlo ke znehodnocení dat hned při jejich tvorbě, a rovněž zvolený styl statistického testování hypotéz vedl pravděpodobně k inflaci hladiny spolehlivosti – dle mého názoru jsou signifikantní výsledky produktem náhodné variability dat. Závěrem pak volba výzkumných otázek hypotéz není v souladu s teorií, statistická úroveň zpracování a reportování je na velmi nízké úrovni, a např. nečekaně opačná souvislost reakčního času a obtížnosti úloh není jakkoli reflektována.

V Brně dne 15. září 2019

Hynek Cígler



**Silné stránky:**

- Teoreticky náročné téma.
- Sehrání scének s amatérskými herci.
- Náročný sběr EEG dat u dětí.
- Pečlivá kontrola intervenujících proměnných ANS trialů (Fibonacciho mřížka).

**Slabé stránky:**

- Chyba ve výpočtu obtížnosti ANS úloh.
- Absence kontroly p-hodnot proti opakovanému testování.
- Stanovení hypotéz ex-post, nerozlišování exploračních a konfirmačních výsledků.
- Slabší provázanost teorie a výzkumného záměru.
- Neodůvodněné vyřazení chybných trialů.
- Chybí interpretace opačného efektu ve vztahu náročnosti úloh a délky jejich řešení.

**Otázky k obhajobě:****Základní otázky:**

1. Jak funguje jednotka obtížnosti ANS úkolů „log“? Jakým způsobem jste ji stanovovala? Srovnajte např. se studií Parka a Brannonové (2014).
2. Zdůvodněte prosím věrohodnost svých výsledků zejm. ve vztahu k chybě I. typu (falešně pozitivní výsledky). Nemohl váš analytický postup vést k inflaci hladiny spolehlivosti? Jaká je dle vás replikovatelnost vašich výsledků?
3. Jak si vysvětlujete, že obtížnější úlohy byly v rozporu s teorií řešené rychleji, a to s velkým efektem –  $r = 0,866$ , resp.  $r = 0,662$  (H7a, H7b)? Jaké jsou možné interpretace a příčiny?

**Doplňující otázky a náměty k diskusi:**

4. Odůvodněte správnost použitého postupu, tedy párového t-testu, při srovnání reakčního času obou skupin (s. 120) vzhledem k interpretaci výsledku.
5. Jaký jev označujete termínem prekognice? V jakém smyslu termín používá vámi citovaný Luck (2014)?
6. Jak velké množství dat bylo z analýz vyřazeno z důvodu chyb v řešení ANS trialu? Jaký mohl být dopad tohoto vyřazení na výsledky statistických testů?
7. Popište prosím podrobně celý statistický a analytický postup, jak byly srovnávány obě skupiny dětí, tedy způsob testování hypotéz H1–H6b.
8. Proč jste vyřadila neúspěšné trial? Jaký to podle vás mohlo mít důsledek pro vaše výsledky?
9. Lze skutečně rozdíl mezi oběma skupinami dětí v reakčním čase (H7c) interpretovat skutečně jako důsledek rozdílné rychlosti ANS systému?

**Reference**

APA (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.