

Univerzita Karlova

Filozofická fakulta

Ústav anglického jazyka a didaktiky

Didaktika konkrétního jazyka

Linda Nepivodová

Teze disertační práce

In Their Own Words and By the Test Results:

*A mixed methods study of the comparison of two
modes of test administration*

Vlastními slovy studentů a podle výsledků testů:

Smíšený výzkum porovnávající dva způsoby administrace testů

Vedoucí práce: Doc. PhDr. Lucie Betáková, MA, Ph.D.

2017

Abstrakt

Tématem disertace je tříletý výzkumný projekt (2014-2016), zabývající se porovnáním povinných testů pokroku z angličtiny, které jsou zadávány univerzitním studentům prvního ročníku katedry anglistiky v České republice ve dvou podobách, a to v tradiční písemné (papírové) formě a formě počítačové. Analýza tří fází projektu, Pilot, Study 1 a Study 2, tvoří ústřední část práce. Hlavní otázka výzkumu spočívá v tom, zda je současná praxe používání počítačové formy testu v tomto kontextu opodstatněná.

Disertace nejprve shrnuje teoretickou debatu a komentuje vybrané podstatné aspekty, důležité pro vlastní srovnání. Zde je aplikována konvergentní paralelní smíšená metoda, kde jsou kvantitativní údaje použity k poměření ekvivalence mezi dvěma způsoby zadávání testu, zatímco kvalitativní údaje reprezentované vlastní zpětnou vazbou jednotlivých studentů osvětlují výhody a nevýhody obou forem testu z jejich úhlu pohledu.

Výsledná interpretace je následující: Výsledky u tradičních písemných i počítačových testů nevykazují ani v jednom stádiu výzkumu žádný statisticky významný rozdíl, ačkoli u papírového testu jsou počty bodů celkově mírně vyšší. Genderové rozdíly nebyly zjištěny, i když se předpokládalo, že studenti mužského pohlaví budou mít lepší skóre u počítačových testů. Významné rozdíly byly pak zjištěny u typu testových otázek. U obou forem testu mají studenti lepší výsledky u volených odpovědí, zatímco u krátkých odpovědí získávají studenti vyšší bodové ohodnocení u písemné formy. U zpětné vazby studentů se aspekt Manipulace významně objevuje mezi nejčastějšími výhodami tradičního písemného testu, který umožňuje dělat si poznámky. Jiné faktory se během výzkumu mění; například Orientaci, která byla uváděna jako jedna z nejčastějších výhod písemné formy u fází Pilot a Study 1, volí jako výhodu u Study 2 jen menšina studentů. Hlavní výhodou počítačové formy testu je pak podle studentů tříletého výzkumu možnost vidět okamžitý výsledek. Zatímco u Study 1 jsou preference ve vztahu k zadané formě testu téměř vyrovnány, u Study 2 preference k počítačové formě již převažují, avšak na rozdíl od Study 1, se kategorie 'Žádná preference' téměř zdvojnásobila. Aplikovaný statistický model zjišťuje, zda existují významné rozdíly mezi výsledkem testu ve vztahu k preferencím studenta. Ty však nebyly nalezeny. V závěru vyjadřuje autorka výzkumu přesvědčení, že bylo shromážděno dostatečné množství důkazů, které v popsaném kontextu plně opravňují používat u testů pokroku počítačovou formu.

Klíčová slova: test pokroku z angličtiny, papírová forma testu, počítačová forma testu, způsoby zadávání, porovnání, konvergentní paralelní smíšená metoda, kvantitativní a kvalitativní data, ekvivalence, vlastní zpětná vazba, gender, typ otázky, manipulace, orientace, okamžitý výsledek, preference.

Abstract

The thesis compares first year university students' compulsory English achievement tests written in two modes of administration, i.e. pencil and paper and computer-based tests (PPT and CBT), over three successive years in an English department in the Czech Republic (2014-2016). The analysis of the three stages of the project, Pilot, Study 1 and Study 2, forms the focus of the study. The central research question investigated is whether the usage of computer-based language testing is justified in such a context.

The dissertation first maps the theoretical discourse and comments on selected fundamental aspects of language testing practices, significant for the comparison itself. A convergent parallel mixed methods design is employed, where quantitative test data is used to measure the equivalency of the two modes and qualitative self-report feedback form data explores the main advantages and disadvantages of these, as viewed by the students examined.

The results demonstrate the following conclusions: The differences in the scores gained from the CBT and PPT modes in all three stages of the research are statistically insignificant, though overall the students perform slightly better in the PPT modes. No striking differences were found concerning gender either, though the prediction put male students' scores in the CBT modes above the female students' ones. Significant differences as regards the question type were then observed. In both modes students perform better with multiple choice questions, while with short answer items students are more successful in the PPT mode. In the students' feedback forms, Manipulation appears to be the most prominent in all the three stages among the advantages of the PPT mode, covering the possibility of making notes. Other factors changed over the time given; Orientation, for example, which in Pilot and Study 1 was among the most frequently cited advantages of the PPT, ceased to be considered as advantageous in Study 2. The main advantage of the CBT mode as identified by the students over the three years is indisputably the immediacy of results. While in Study 1 the preferences concerning the two modes are almost equally divided, in Study 2, the CBT is preferred to the PPT, however, the category of 'No preference' almost doubled. The statistical model is employed to ascertain whether there are statistically significant differences between the mean subtest scores in connection to student preferences, however, none were found. As concluded, the researcher believes that sufficient evidence has been provided to consider the usage of computer-based achievement tests in the context given to be justified.

Key words: English achievement test, pencil and paper-based test (PPT), computer-based test (CBT), modes of administration, comparison, convergent parallel mixed methods design, quantitative and qualitative data, equivalency, self-report feedback, gender, question type, manipulation, orientation, immediacy of results, preference.

TEZE DISERTAČNÍ PRÁCE

1. Tematické vymezení

Disertační práce se zaměřuje na porovnávání počítačových a papírových verzí testů z anglického jazyka používaných při testování studentů prvního ročníku bakalářského studijního programu Anglický jazyk a literatura na Filozofické fakultě Masarykovy Univerzity v Brně. Jedná se o testy, které mapují pokrok studentů v rámci semestrů, tzv. achievement tests neboli testy pokroku.

Práce navazuje na výzkum, který jsem provedla v rámci magisterské práce a jehož cílem bylo porovnat papírové a počítačové verze postupových zkoušek jazykové úrovně, tzv. proficiency exams, jak z hlediska obsahu, tak způsobu administrace. Tato studie vykazala pozoruhodné výsledky, zároveň však přinesla mnoho otázek a poukázala na určité problémy. Na ty jsem se v předkládané práci doktorské zaměřila, výzkum rozšířila z původních 36 studentů v magisterské práci na 378 a snažila se proniknout co nehlouběji do problematiky počítačového jazykového testování.

Hlavní teoretické principy, kterými se práce zabývá jsou:

- Zasazení celé problematiky do historického vývoje testování
- Fundamentální koncepty jazykového testování, například validita, reliabilita, washback, praktikalita
- Dichotomické vztahy různých typů testů relevantních pro kontext práce, například testy ověřující versus rozlišující, testy integrativní versus zaměřené na jednu doménu, formativní versus sumativní hodnocení, atd.
- Rámec pro počítačové testování (Suvorov a Hegelheimer, 2014) rozdělený do 9 částí, z nichž každá konkrétně dokumentuje hlediska známých odborníků v této oblasti a tyto dále zpracovává v diskusi
- Počítačové testování – výhody a nevýhody podle J.D.Browna (1997), Chapelle and Douglase (2006), Noyes and Garlanda (2008), H. D. Browna (2010) a dalších

(viz Příloha, kde jsou některé principy detailněji rozpracovány.)

2. Hlavní zkoumané problémy

Obecným cílem výzkumu je posoudit, do jaké míry jsou počítačové verze testů pokroku používané na Katedře anglistiky FF v Brně porovnatelné s tradičními testy papírovými.

Z tohoto cíle je odvozena formulace výzkumných otázek:

Hlavní výzkumná otázka:

Je používání počítačových testů pokroku v kontextu českého terciálního vzdělávání studentů angličtiny prvního ročníku opodstatněné?

Hlavní výzkumná otázka se dále dělí na kvantitativní, kvalitativní, a podle Tashakkoriho a Creswella (2007, p.208) integrované či hybridní pod-otázky:

1. Existují ve výsledcích z počítačového a papírového způsobu administrace testů nějaké významné rozdíly?
2. Liší se výsledky z počítačového a papírového způsobu administrace testů v závislosti na pohlaví?
3. Liší se výsledky z počítačového a papírového způsobu administrace testů s ohledem na typ testové otázky (volená odpověď versus krátká tvořená odpověď)?
4. Jaké jsou výhody a nevýhody papírového a počítačového způsobu administrace testů z pohledu studentů?
5. Jaké jsou preference studentů ohledně papírového a počítačového způsobu administrace testů?
6. Je nějaká souvislost mezi preferencemi studentů a jejich dosaženými výsledky?

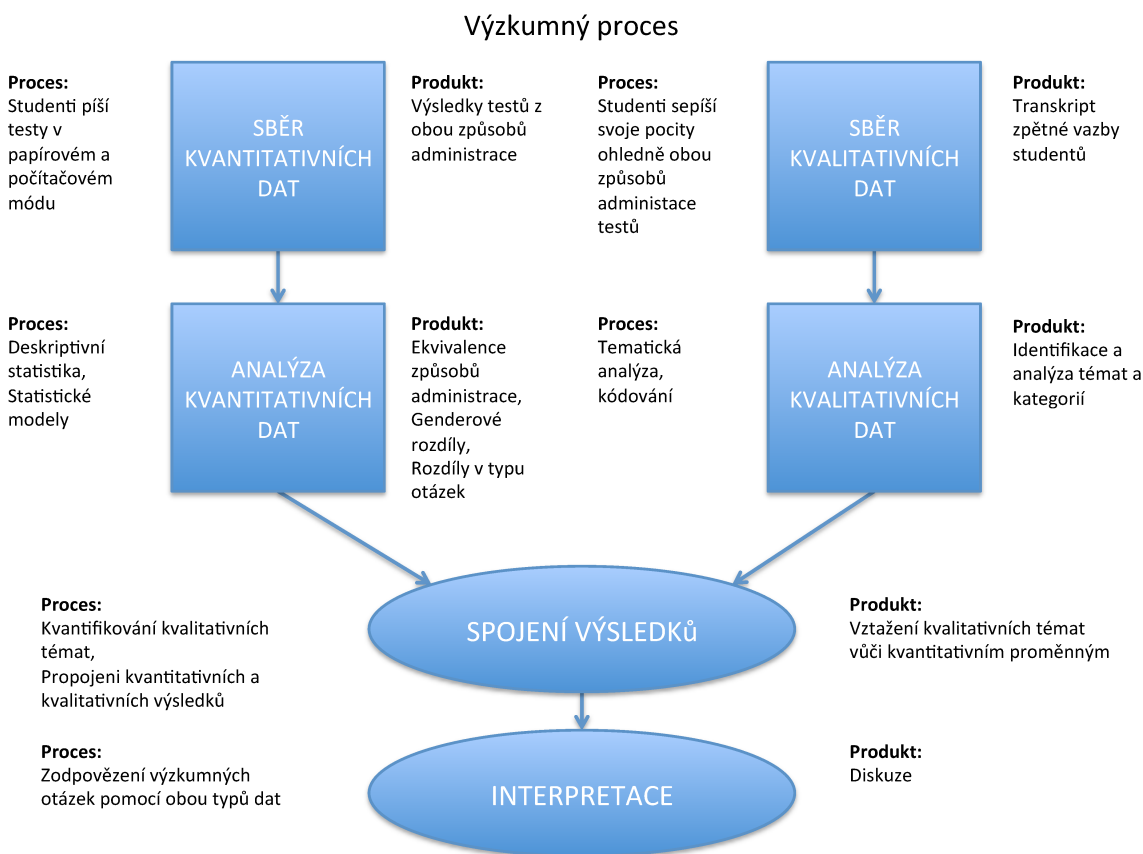
První tři otázky se týkají analýzy kvantitativních dat, otázky 4 a 5 jsou zodpovězeny pomocí dat kvalitativních a poslední otázka propojuje ve výsledné interpretaci oba typy dat.

Praktickým výstupem disertační práce by mělo být navržení výsledné účinné metody testování pro studenty bakalářského studijního programu Katedry anglistiky.

3. Struktura

Práce je rozdělena do dvou částí. V první části jsou uvedena teoretická východiska, představen stav poznání a celá problematika je dokumentována předešlými výzkumy v této oblasti. Druhá část se pak zabývá vlastním výzkumem, jeho metodikou, realizací, výsledky a jejich rozbořem.

4. Diagram procesu výzkumu



4. Výzkumný vzorek

Základní soubor tvoří studenti prvních ročníků bakalářského studia Anglického jazyka a literatury na FF MU v Brně v letech 2014-2016. Podle Soukupa a Kočvarové (2016) by se u kvantitativní analýzy mělo pracovat s celou populací, pokud se celkový počet pohybuje kolem 100 účastníků výzkumu. V tomto případě se v každé fázi výzkumu (Pilot, Study 1 a Study 2), které na sobě byly nezávislé a každé fáze se účastnili jiní studenti, pracovalo s celou populací denních studentů prvního ročníku, tedy cca se 120 studenty. Protože jsem chtěla získat co nejkomplexnější informace popisující daný výzkumný problém,

nechtěla jsem počet studentů zmenšovat ani v následném kvalitativním sběru dat, pak by totiž bylo velmi obtížné data propojovat a integrovat.

5. Metody

Metodu výzkumu je možné charakterizovat jako smíšenou paralelně konvergentní, kde jsou jak kvantitativní, tak kvalitativní údaje, analyzované nejprve odděleně, aby byly později spojeny za účelem výsledné interpretace, jak bylo popsáno v diagramu výše.

Kvantitativní část:

Studenti v každé fázi výzkumu skládají test pokroku v obou způsobech administrace (papírové i počítačové). Pouze v Pilotu nejprve píší všichni papírovou, a pak teprve počítačovou verzi testu. Ve Study 1 i 2 je již využita technika vyvažování (counterbalancing), aby se zamezilo ovlivňování výsledků testů způsobeného pořadím. Jedná se o experimentální design výzkumu, a protože jsou všichni studenti vystaveni oběma experimentálními podmínkami, jsou si zároveň kontrolní skupinou (Johnson a Christensen, 2008, p.320). Výsledky jsou pak analyzovány.

Kvalitativní část:

Studenti po každém testu ihned podají zpětnou vazbu na mód testu, který právě složili, vyplněním formuláře, který obsahuje dvě až čtyři otevřené otázky (podle fáze výzkumu). Vyjadřují se k pozitivům a negativům daného způsobu administrace z jejich úhlu pohledu, uvádějí své preference, apod. Jejich odpovědi jsou transkribovány, kódovány otevřeným způsobem kódování a analyzovány. Na základě jejich odpovědí jsem vytvořila také dotazník, který má za sebou zatím pilotáž a předběžnou analýzu, ale jakmile bude dokončen a validován, bude sloužit k zjišťování názorů a pocitů velkého počtu studentů časově efektivním způsobem.

6. Závěry

1. Existují ve výsledcích z počítačového a papírového způsobu administrace testů nějaké významné rozdíly?

Kvantitativní analýza výsledků z testů ukázala, že mezi jednotlivými způsoby administrace nejsou statisticky významné rozdíly, oba módy tak mohou být používány zaměnitelně. Zajímavostí je, že ve všech třech fázích výzkumu (tzn.

Pilot, Study 1 a Study 2), studenti celkově dosáhli mírně lepších výsledků v papírovém módu administrace. Tyto výstupy se však shodují se studii popsány v teoretické části práce, například, meta-analýzou Meada a Drasgowa, která se zabývala 28 studii porovnávajícími dva způsoby administrace testů.

2. Liší se výsledky z počítačového a papírového způsobu administrace testů v závislosti na pohlaví?

Na základě poznatků prodiskutovaných v teoretické části byl předpoklad, že muži budou dosahovat lepších výsledků v počítačovém módu testu než ženy. Toto se však statisticky neprokázalo. Možný vliv pohlaví bude dále popsán v návaznosti na preference ve výzkumné otázce č. 5.

3. Liší se výsledky z počítačového a papírového způsobu administrace testů s ohledem na typ testové otázky (volená odpověď versus krátká tvořená odpověď)?

Této otázce byla věnována pozornost ve Study 2 na základě smíšených výsledků ze Study 1 a byly zjištěny následující statisticky významné rozdíly. Studenti v obou módech dosahují celkově lepších výsledků u otázek s volenou odpovědí, což se dá očekávat, jelikož tyto otázky testují pouze receptivní dovednosti, nikoli produkci jako je tomu v případě odpovědí tvořených. U krátkých tvořených odpovědí však studenti mají statisticky signifikantně lepší výsledky v papírovém módu na rozdíl od módu počítačového.

4. Jaké jsou výhody a nevýhody papírového a počítačového způsobu administrace testů z pohledu studentů?

Mezi nejčastější pozitiva papírového módu testu patří podle studentů Manipulace, Orientace, Žádná Technika, Čas a Postoj. Ohledně Manipulace, studenti nejčastěji zmiňují možnost dělat si poznámky, podtrhovat, zaškrtnout, apod. Někteří jsou přesvědčeni, že se jim tímto způsobem lépe přemýšlí o odpovědích. Lepší Orientaci či navigaci v papírové verzi testu spojují hlavně s možností vidět test jako celek. Téma Žádná Technika je reflektováno i v negativech testů počítačových, kde naopak Technika představuje jeden

z hlavních problémů. Studenti si stěžují na problémy s přihlašováním, rušivé zvuky klávesnice ostatních, obrazovku, a technické problémy obecně. Téma Čas zahrnuje dva hlavní aspekty – trvání testu a přítomnost/ absenci měření času. Zatímco v papírové verzi studenti považují test celkově za rychlejší, v počítačové verzi oceňují možnost kontrolovat si, kolik mají času pomocí časoměřiče. Postoje jsou jedním z hlavních témat v Pilotu, kde jsou studenti u papírové verze testu klidnější a méně nervózní, ovšem ve Study 1 a Study 2 se již toto téma nevyskytuje tak často. Co se týká nevýhod počítačového módu, studenti nejčastěji přicházejí s odpovědí „Nic“, což se dá interpretovat i jako výhoda, a dále pak s výše zmíněnými problémy s Technikou, nebo s komentáři ohledně Obsahu a Času.

Za hlavní výhody počítačové verze testu pak studenti považují okamžité Výsledky (jejichž absence je zároveň nejčastěji citovanou nevýhodou módu papírového), možnost snadno Měnit odpovědi, a také Čas ve spojitosti s časoměřičem. Mezi nevýhodami papírové verze převažuje odpověď „Nic“ a výše zmíněná absence Výsledků. Někteří studenti si také stěžují na písmo a obávají se, že nebude čitelné, nebo celkovou neupravenost při Měnění odpovědí. V jejich odpovědích je patrný posun od negativnějšího vnímání počítačové verze během Pilotu k vnímání pozitivnějšímu ve Study 2.

5. Jaké jsou preference studentů ohledně papírového a počítačového způsobu administrace testů?

Posun preferencí papírové verze testu k verzi počítačové, zkoumaných ve Study 1 a Study 2, je znatelný. Pokud jde o genderové rozdíly, je jasně patrné, že již ve Study 1 muži mírně preferují počítačovou verzi a ženy papírovou. Ve Study 2 již obě pohlaví preferují počítačovou verzi před papírovou, i když nejpočetnější kategorií se stává „Žádná preference“. Možné vysvětlení je, že studenti si již navykli na počítačové testování a na módu administrace jim tolik nezáleží.

6. Je nějaká souvislost mezi preferencemi studentů a jejich dosaženými výsledky?

Statisticky významný vliv preferencí na výsledky studentů nebyl prokázán. Zajímavé je, že bylo více případů, kdy studenti preferovali papírovou verzi a opravdu v ní dosáhli i lepších výsledků, ale toto může být dáno tím, že většina

studentů měla papírové skóre lepší než počítačové, i když rozdíly byly zanedbatelné.

Hlavní výzkumná otázka:

Je používání počítačových testů pokroku v kontextu českého terciálního vzdělávání studentů angličtiny prvního ročníku opodstatněné?

Na základě předložených výsledků výzkumu se domnívám, že používání počítačového způsobu administrace testu pokroku v uvedeném kontextu je opodstatněné. Mírné rozdíly ve výsledcích obou módů nebyly statisticky významné a postoje a pocity studentů se staly pozitivnějšími vzhledem k počítačovému způsobu administrace.

7. Limitace výzkumu

- Design výzkumu byl v průběhu modifikován, jelikož jsem se snažila vždy vylepšit, co v předcházející fázi nezafungovalo. Zatímco je tento postup vítaný a běžný v kvalitativním výzkumu, v kvantitativním může být kritizován. Vzhledem k faktu, že výzkum je smíšený a dlouhodobý a tři fáze (Pilot, Study 1 a Study 2) na sobě nejsou závislé, s modifikací by neměl být problém.
- Hloubka kvalitativních dat byla obětována na úkor zachování co největšího počtu respondentů, a získání tak celistvého obrazu. Z tohoto důvodu byly použity formuláře s otevřenými otázkami, místo např. hloubkových rozhovorů s extrémními případy.
- Šesti-týdenní pauza mezi testem pokroku 1 a 2 v Pilotu a ve Study 1 může být považována za úskalí, protože do procesu vstupuje mnoho proměnných - za danou dobu se mohou změnit jak dovednosti, tak čas, který studenti studiu věnují, apod. To je jeden z hlavních důvodů, proč byl design výzkumu změněn ve Study 2 a testy v obou módech proběhly v jeden den.
- Věk účastníků výzkumu je rozhodně důležitým faktorem a výběr pouze studentů prvního ročníku může být vnímán jako limitující.
- Do výzkumu jsem chtěla zahrnout i postoje učitelů vůči počítačovému způsobu administrace, jelikož se domnívám, že mohou velice ovlivnit postoje studentů, ale vzhledem k překročení rozsahu práce jsem se rozhodla je do výzkumu již nezařadit.

8. Zdůvodnění významu výzkumu

Jedná se o první studii takového rozsahu, zabývající se porovnáním papírového a počítačového způsobu administrace jazykových testů pokroku v univerzitním kontextu. Přináší analýzu nejen kvantitativních, ale i kvalitativních dat, která přinášejí nové poznatky přímo od studentů, a mapují tak hlavní výhody a nevýhody počítačového testování z úhlu pohledu studentů, nikoli teoretiků, testerů či administrátorů testů. Zejména kvalitativní data, pocházející od více než 350 studentů, mají i externí validitu a rozhodně mohou být využity jako užitečný vhled do problematiky počítačového testování pro instituce, které uvažují nad přechodem k tomuto módu testování. Domnívám se, že jazykové testování obecně představuje velmi opomíjenou oblast v ČR a doufám, že se současnou disertací podaří poukázat, jak na nutnost začít se testováním více zabývat, tak na problematické aspekty přechodu z papírového na počítačové testování, které právě často probíhá bez většího odborného přístupu k dané problematice.

9. Další výzkum

Další výzkum je přímo navázán na limitace výzkumu zmíněné výše:

- Získat detailnější data, například pomocí hloubkových rozhovorů s extrémními případy (např. jejichž rozdíly ve výsledcích u obou způsobů administrace se nejvíce liší)
- Navrhnout a provést specifitější design výzkumu. Po zdokumentování ekvivalence módů v daném kontextu bude možné zabývat se určitými aspekty počítačového testování, například rozdíly dané různým typem otázek, možnostmi dělat si poznámky, funkcí časoměřiče, atd.
- Zahrnout do výzkumu jinou věkovou kategorii, například kombinované studenty, kteří, alespoň v našem kontextu, bývají starší a počítačové testování pro ně může být problematičtější.
- Vliv postojů učitelů k počítačovému testování na postoje studentů.
- Validace dotazníku vycházející z kvalitativních dat získaných v tomto výzkumném projektu.

- Důležitost a potřeba replikačních studií. Z tohoto důvodu jsem se také rozhodla poskytnout všechna data včetně statistických modelů, kódování, atd. online se záměrem usnadnit postup dalším badatelům v této oblasti.

10. Vybraná literatura

- Alderson, J. C. (2000). Technology in testing: The present and the future. *System*, 28(4), 593-603.
- Alderson, J. C. et al. (1995). *Language test construction and evaluation*. New York, USA: Cambridge University Press.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford: Oxford University Press.
- Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01.
- Beckers, J. J., & Schmidt, H. G. (2003). Computer experience and computer anxiety. *Computers in Human Behavior*, 19(6), 785-797.
- Bennett, R. E. (2011). Formative assessment: A Critical review. *Assessment in Education: Principles, Policy and Practice*, 18(1): 5-25.
- Black, P.J. & Wiliam, D. (2012). Assessment for learning in the classroom. In Gardner, J. (ed.) *Assessment and learning*. 2nd ed. pp. 14-44, London: Sage.
- Broos, A. (2005). Gender and information and communication technologies (ICT) anxiety: Male self-assurance and female hesitation. *CyberPsychology & Behavior*, 8(1), 21-31.
- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: principles and classroom practices*. Upper Saddle River, N.J: Pearson Education.
- Brown, J. D. (1997). Computers in language testing: Present research and some future directions. *Language Learning & Technology*, 1(1), 44-59.
- Brown, J. D. (2005). *Testing in language programs: A comprehensive guide to English language assessment*. Upper Saddle River, N.J.: Prentice Hall Regents.
- Carr, N. T. (2011). *Designing and analyzing language tests*. Oxford: Oxford University Press.
- Chapelle, C. A., & Douglas, D. (2006). *Assessing language through computer technology*. Cambridge, England: Cambridge University Press.
- Creswell, J. W., & Clark, V. L.P. (2011). *Designing and conducting mixed methods research*. Thousand Oaks: Sage.
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*. (4th ed.) Los Angeles: Sage.
- Davey, T. (2011). *Practical considerations in computer-based testing*. Princeton, NJ: Educational Testing Service.
- Dudeny, G. Hockly, N & Pegrum, M. (2013). *Digital literacies*. Malaysia: Pearson.

- Fulcher, G. (2010). *Practical language testing*. London: Hodder Education.
- Gardner, J.(ed.), (2012). *Assessment and learning*. Los Angeles, Calif: Sage.
- Green, A. (2014). *Exploring language assessment and testing: Language in Action*. Abingdon, Oxon: Routledge.
- Hughes, A. (1989, 2013). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Jamieson, J. (2005). Trends in computer-based second language assessment. *Annual Review of Applied Linguistics*, 25, 228-242.
- Joiner, R., Gavin, J., Brosnan, M., Cromby, J., Gregory, H., Guiller, J., Maras, P., Moon, A. (2013). Comparing first and second generation digital natives' Internet use, Internet anxiety, and Internet identification. *Cyberpsychology, Behavior and Social Networking*, 16 (7), 549-52.
- Lightstone, K., & Smith, S. M. (2009). Student choice between computer and traditional paper-and-pencil university tests: what predicts preference and performance?. *Revue internationale des technologies en pédagogie universitaire/International Journal of Technologies in Higher Education*, 6(1), 30-45.
- Maxwell, J. A. (2013). *Qualitative research design: An interactive approach*. Thousand Oaks, Calif: SAGE Publications.
- Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, 114(3), 449-458.
- Morrow, K. (2012). Communicative language testing. In Coombe, C. A., In Davidson, P., In O'Sullivan, B., & In Stoyhoff, S. *The Cambridge guide to second language assessment*. pp.140-146, NY: Cambridge University Press.
- Nepivodová, L. (2006). On communicative language competence, validity and different modes of administration. MA thesis, Department of English and Americal Studies, Masaryk University.
- Noyes, J. M., & Garland, K. J. (2008). Computer-vs. paper-based tasks: Are they equivalent?. *Ergonomics*, 51(9), 1352-1375.
- Ockey, G. J. (2009). Developments and challenges in the use of computer-based testing for assessing second language ability. *The Modern Language Journal*, 93, 836-47.
- Paek, P. (2005). Recent trends in comparability studies. *Pearson Educational Measurement*. Retrieved from https://www.researchgate.net/profile/Pamela_Paek/publication/245023911_Recent_Trends_in_Comparability_Studies_Using_testing_and_assessment_to_promote_learning/links/00b7d51d5c29b537b5000000.pdf
- Soukup, P., & Kocvarová, I. (2016). Velikost a reprezentativita výberového souboru v kvantitativne orientovaném pedagogickém výzkumu 1. *Pedagogická orientace*, 26(3), 512-536.
- Suvorov, R., & Hegelheimer, V. (2014). *Computer-assisted language testing*. The companion to language assessment. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/9781118411360.wbcla083/full>
- Švaříček, R., & Šed'ová, K. et al. (2007). *Kvalitativní výzkum v pedagogických vědách*. Praha: Portal, s.r.o.
- Tapscott, D. (2009). *Grown up digital: How the net generation is changing your world*. New York: McGraw-Hill.
- Wang, S., Jiao, H., Young, M. J., Brooks, T., & Olson, J. (2008). Comparability of computer-based and paper-and-pencil

testing in K–12 reading assessments: A meta-analysis of testing mode effects. *Educational and Psychological Measurement*, 68(1), 5-24.

Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Basingstoke: Palgrave Macmillan.

Weir, C., Yan, J., O'Sullivan, B., & Bax, S. (2007). Does the computer make a difference?: The reaction of candidates to a computer-based versus a traditional hand-written form of the IELTS Writing component: effects and impact. *International English Language Testing System (IELTS) Research Reports 2007: Volume 7, 1*.

White, D.S. and Le Cornu, A. (2011). Visitors and residents: A new typology for online engagement. *First Monday*, 16(9), Retrieved from <http://firstmonday.org/article/view/3171/3049>.

Whitley, B. E. (1997). Gender differences in computer-related attitudes and behavior: A meta-analysis. *Computers in Human Behavior*, 13(1), 1-22.

Winke, P., & Fei, F. (2008). Computer-assisted language assessment. In N. Van Deusen-Scholl & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (Vol. 4, pp. 353–64). New York, NY: Springer.

11. Odborná a publikační činnost

2016/08/24 - 2016/08/24: Masaryk University, Brno, CZE

Teacher Training courses for Fulbright English Teaching Assistants

2015/08/26 - 2016/08/28: Masaryk University, Brno, CZE

Teacher Training courses for Fulbright English Teaching Assistants

2014/09/07 – 2014/09/13: University of Szeged, Szeged, HUN

Guest Lectures in Language Testing and Teacher training for MA students of
Department of Applied Linguistics

2014/08/27 - 2016/08/29: Masaryk University, Brno, CZE

Teacher Training courses for Fulbright English Teaching Assistants

2013/08/02 – 2013/11/29: University of Cincinnati, Cincinnati, Fulbright Grants,
USA

Visiting scholar, research stay

2012/09/03 – 2012/09/07: University of Cambridge, Cambridge, GBR

Participation at Language Testing course taught by Prof Cyril Weir and Dr Lynda
Taylor

2012/08/06 – 2012/08/10: Göteborg University, Göteborg, SWE

Participation at Language Testing: A Psychometric Perspective

2012/12/03 – 2012/12/07: University of Bristol, Bristol, GBR

Guest Lectures and Seminars in Language Testing for bachelor degree students
studying Slavonic Languages

2011/03/13 – 2011/03/17: National and Kapodistrian University of Athens,
Athens, GRC

Guest lectures in Language Testing and ICT for MA students studying Teaching
English as a Foreign Language

Cambridge Let's Speak NEPIVODOVÁ, Linda. Cambridge Let's Speak. In *British Council Cambridge Let's Speak Project*. 2016.

Online testování v ELFu: Metodika a praktické postupy NEPIVODOVÁ, Linda. Online testování v ELFu: Metodika a praktické postupy. In *ELF day 2016*. 2016.

Are Students and Teachers Ready for Computer-based Language Exams? A Comparison of Test Mode Delivery NEPIVODOVÁ, Linda. In *12th EALTA Conference 2015, University of Copenhagen, Copenhagen, Denmark*. 2015.

A tale of two paradigms: a comparison of test mode delivery on either side of the Atlantic NEPIVODOVÁ, Linda. In *11th EALTA Conference Warwick, Warwick, United Kingdom*. 2014.

Interkulturní komunikační kompetence v angličtině by Zerzová, J. Recenze by NEPIVODOVÁ, Linda. In *Universitas, 4, 2013. ISSN 1211-3387*.

Two Worlds Collide NEPIVODOVÁ, Linda. Pencil and paper versus computer-based tests. In *10th EALTA conference, Istanbul*. 2013.

Comparing two modes of test administration NEPIVODOVÁ, Linda. Comparing two modes of test administration. In *9th EALTA Conference Innsbruck, Austria*. 2012.

SIG for Assessing for Academic purposes member NEPIVODOVÁ, Linda. *SIG member*. : EALTA, 2012.

EALTA member NEPIVODOVÁ, Linda. *EALTA member*. : EALTA, 2012.

Chci studovat angličtinu VANDERZIEL, Jeffrey Alan, Linda NEPIVODOVÁ, Lidia KYZLINKOVÁ, Jarmila FICTUMOVÁ a Věra VÉMOLOVÁ. *Chci studovat angličtinu*. 2. vyd. Brno: Barrister & Principal, 2012. 183 s. ISBN 978-80-87474-43-3.

Research Activities

Computer-based Language Testing, The Role of Validity in Language Testing, Test

Development, Entrance Exams at Tertiary Level, Methodology, Acquisition vs Learning, Error

Correction

Responsibilities and Courses taught at home institution:

Responsibilities: Entrance exams development and analysis, in charge of language exams administered by the Department of English and American Studies, responsible for Mock Exams, final theses supervision

Courses taught: Language Testing, Intensive Methodology Course, Language Acquisition, Introduction to Research Methods, Internal and External Teaching Practice, Micro-teaching and Observation, and Practical English.

Theses supervised:

Implementation of E-learning in Homework for Secondary School Learners

diplomová práce, Mgr. et Mgr. Dana Pětrošová, 17. 6. 2016

The Promotion of Study Habits and Learner Autonomy at Czech Secondary Schools

diplomová práce, Mgr. Magdaléna Marková, 17. 6. 2016

Kinestetický způsob učení Angličtiny na vyšším stupni základních škol a na středních školách

diplomová práce, Mgr. Ondřej Svoboda, 16. 6. 2016

Error Correction: Qualified and Non-qualified Native English Speaking Teachers and Students' Perceptions

diplomová práce, Mgr. Silvie Vinklerová, 16. 6. 2016

Teaching Politeness in Company Business English Courses

diplomová práce, Mgr. Zuzana Danielová, 16. 6. 2016

Praise and Feedback in English Classes at Czech Secondary Schools

diplomová práce, Mgr. Kristina Halfarová, 15. 6. 2016

Teaching English in Third World Countries - Teacher's Perspective

bakalářská práce, Bc. Tereza Šmilauerová, 14. 6. 2016

Error Correction in Second Language Learning

bakalářská práce, Bc. Mario Cabal, 27. 1. 2016

[Graded Readers as a Tool to Raise Learners' Interest in Reading](#)

diplomová práce, Mgr. Daniela Kleinová, 25. 1. 2016

[The use of Interactive Whiteboards in English lessons at Czech schools](#)

diplomová práce, Mgr. Kateřina Štruncová, 25. 1. 2016

[Foreign Language Anxiety in Speaking Activities: Students' Perspectives](#)

diplomová práce, Mgr. Aneta Krůpová, 19. 6. 2015

[Influence of Selected Motivational Strategies on English Learners in Company Courses](#)

diplomová práce, Mgr. Hana Humpolíková, 19. 6. 2015

[Using Outdoor Environment in English Language Teaching](#)

diplomová práce, Mgr. Kateřina Janáková, 19. 6. 2015

[Self-Assessment in English Lessons at Grammar Schools](#)

diplomová práce, Mgr. Tereza Heisigová, 19. 6. 2015

[English Language Proficiency in Spain: Situation in Secondary Schools](#)

bakalářská práce, Bc. Petra Seidlová, 28. 1. 2015

[Harry Potter Teaches English: An English Summer Camp Design](#)

diplomová práce, Mgr. Lucie Svobodová, 10. 6. 2014

[The Relationship Between Mind Sets and the Motivation of Secondary School Students of English as a Second Language](#)

diplomová práce, Mgr. Eva Čepelková, 28. 1. 2014

[Comparison of Adult EFL Textbooks on Pre-Intermediate Level](#)

diplomová práce, Mgr. Pavlína Švejdová, 14. 6. 2013

[Global Development Education and Czech ELT Context](#)

diplomová práce, Mgr. Anna Halfarová, 13. 6. 2013

[Teaching Literature in Secondary Schools](#)

diplomová práce, Mgr. Eva Mlčáková, 13. 6. 2013

[Error Correction in Spoken Practice](#)

diplomová práce, Mgr. Gabriela Tomková, 12. 6. 2013

Theses reviewed:

[The Most Difficult Aspects of Learning The English Language for Grammar Schools Students](#)

bakalářská práce, Bc. Jan Balek, 15. 6. 2015

[Teaching English through Literature: Course Design](#)

diplomová práce, Mgr. et Mgr. Pavla Nováková, 29. 1. 2015

[Errors and Error Treatment in Secondary School Student Writing](#)

diplomová práce, Mgr. Petra Kohoutová, 10. 6. 2014

[Foreign language spoken performance anxiety](#)

diplomová práce, Mgr. Michaela Slezák Polónyová, 9. 6. 2014

[First Language Acquisition: A Comparative Study of English and Hungarian](#)

diplomová práce, Mgr. Veronika Izsáková, 9. 6. 2014

[Internet Slang in Relation to ELT](#)

diplomová práce, Mgr. et Mgr. Kristýna Guricová, 27. 1. 2014

[Teaching and Learning English Verbs Effectively](#)

diplomová práce, Mgr. Pavla Kadová, 14. 6. 2013

[Cat in the Rain: The suitability of an authentic literary work for the intermediate English classroom](#)

diplomová práce, Mgr. Hana Tichá, 13. 6. 2013

[New tendencies in RP: Czech secondary school materials](#)

diplomová práce, Mgr. Alexandra Carreño Pořízová, 29. 1. 2012

12. Příloha: Vybrané teoretické principy

1.1 VALIDITA

- Validita, neboli platnost testu se snaží zodpovědět na tu nejdůležitější otázku v jazykovém testování, a to zda test skutečně testuje to, co má v úmyslu testovat. (Hughes, Alderson, Henning, Weir, et al.)
- Alderson poukazuje na fakt, že jedním z nejčastějších problémů je zneužití testu (test misuse), ke kterému dochází v momentě, kdy je test pro určitý účel použit k účelu jinému.
- Pokud test není validní k dosažení účelů, pro které byl vytvořen, výsledky z takového testu nemají vypovídající hodnotu.
- Pojem validita představuje obrovský koncept, který doprovázejí určité kontroverze a polemiky.
- Nejznámější pohled na validitu jako jednoduší koncept nabídl Messick. S ním souhlasil Bachman, ale proti nim se postavili Alderson a Weir, kteří podporují mnohostrannost tohoto termínu.

1.1.1 VALIDITA OBSAHU (Content validity)

- Podle Hughese má test validitu obsahu, pokud je jeho obsah složen z reprezentativního vzorku jazykových kompetencí, které má test ověřit.
- Porovnáváním specifikací testu a jeho obsahu se dá validita obsahu určit.
- Weir se přiklání k pojmenování Kontextová validita.

1.1.2 VALIDITA VZHLEDU (Face Validity)

- Tento aspekt testu, který je velice často považován za nevědecký a zanedbatelný, si podle mého názoru zaslouží pozornost, protože i samotný vzhled testu velice ovlivňuje celý testovací proces.
- Validita vzhledu také ovlivňuje důvěryhodnost a přijatelnost testu pro veřejnost.
- Podle Hughese, test má validitu vzhledu pokud vypadá, že testuje to, co má testovat.
- Cattell (1964), Cronbach (1984) et al. považují validitu vzhledu za nepotřebnou.

1.1.3 VALIDITA ODPOVĚDÍ (Response Validity)

- Jak název napovídá, validita odpovědí se zabývá odpověďmi testovaných subjektů a to zejména tím, jak se k dané úloze postaví, jaké procesy a postup v odpovídání zapojí.
- Je úzce spjata a ovlivněna formátem odpovědí.
- Např. Hughes vypočítává nevýhody používání Multiple Choice Questions (otázky s volenou odpovědí): testuje pouze receptivní dovednost (oproti produkci), hádání ovlivňuje výsledek, má omezené možnosti vzhledem k tomu, co testovat, je těžké napsat úspěšné položky s validními distraktory, špatný washback, atd.

1.1.4 VALIDITA SOUVISEJÍCÍ S KRITÉRII (Criterion-related Validity)

- Validita určující, jak se výsledky jednotlivce vztahují k nějakému jinému nezávislému hodnocení.
- Alderson ji nazývá Externí validitou (1995).
- Dále se rozděluje na validitu SOUBĚŽNOU (i.e. jak výsledky korelují s nějakým jiným testem v podobném čase) a validitu PREDIKČNÍ (i.e. jak mohou výsledky předpovědět budoucí projev).

1.1.5 VALIDITA BODOVÁNÍ (Scoring Validity)

- V literatuře tuzemské i zahraniční obvykle nazývána RELIABILITOU – např. Hughes, Alderson, Anastasi, Dvořáková)
- Bachman a Palmer ji definují jako spolehlivost hodnocení, kdy podle nich spolehlivý test bude vykazovat stejné hodnoty napříč různými specifiky dané testovací situace.
- Já se přikláním k termínu Scoring Validity, jak kvůli zmírnění tenze mezi VALIDITOU A REALIABILITOU (čím je test validnější, tím je jeho reliabilita horší a naopak), tak kvůli tomu, že VALIDITA by měla být primárním cílem a její součástí, nikoli protipólem, by mělo být spolehlivé hodnocení.
- 3 hlavní typy zjišťování spolehlivosti testů – Test/Retest (jeden test administrován dvakrát po sobě), Reliabilita paralelních forem (dvě paralelní verze testu), Metoda Split half (v rámci jednoho testu, např. sudé, liché položky - testuje se vnitřní konzistence).

- Také rozlišujeme Inter (shodu mezi různými hodnotiteli) a Intra (shodu v rámci konzistence hodnocení u jednoho hodnotitele) spolehlivost raterů.

1.1.6 KONSTRUKTOVÁ VALIDITA neboli VALIDITA CELKU (Construct Validity)

- Jeden z nejkomplicovanějších konceptů v celém procesu vývoje jazykových testů, např. Messick považuje Konstruktovou validitu za nadřazenou všem ostatním druhům validity.
- Je úzce spjata s teorií konstruktů, Hughes tvrdí, že test má konstruktovou validitu pokud se dá dokázat, že měří přesně ty schopnosti, které ten daný teoretický konstrukt má měřit.
- Velká debata ohledně toho, co je vlastně konstrukt. (Bachman, Cronbach, Weir, et al.)

1.1.7 OTÁZKY OHLEDNĚ VALIDITY

- Validita testu je ovlivňována nedostatky testu. Nízkou validitu budou mít za následek nejasné a nesrozumitelné pokyny, příliš obtížný jazyk testu, špatně diskriminující položky (příliš snadné či obtížné), nejednoznačné položky, atd.
- Nízkou validitu také způsobuje špatná administrace, např. opisování žáků, či špatné vyhodnocení (např. při volně tvořených odpovědích – pokud nedojde k moderaci klíče – různí hodnotitelé budou hodnotit různě).
- Validita také může být snižována faktory na straně studenta – např. nervozita, strach, atd.

1.2 WASHBACK

- Také nazýván backwash, neboli dopad testu na studenta, proces učení, třídu, atd.
- Může být pozitivní nebo negativní.
- Jedna z nejvíce diskutovaných oblastí počítačového testování – jaký má počítačové testování vlastně dopad na studenty?
- Chapelle a Douglas jsou velice kritičtí a domnívají se, jak bylo zmíněno výše, že počítačové testování může mít negativní dopad na studenta, učení, atd.
- Pro studenty v počítačích nepříliš zdatné mohou počítačové testy vyvolat paniku a stres.

1.3 POČÍTAČOVÉ TESTOVÁNÍ

- Implementace počítačů jak do výuky, tak do testování jde ruku v ruce s rychlým vývojem moderních technologií.
- Počítačové testy (pokud to vybavenost školy dovolí) často nahrazují tradiční papírové, aniž se zjišťuje, zda jsou tyto dva odlišné způsoby administrace testu porovnatelné.

1.3.1 ROZDÍLY V METODÁCH POČÍTAČOVÉHO TESTOVÁNÍ

- Místní a časové rozdíly – test může být zadáván kdekoli, kde je přístup na internet, bez dozoru, rezervace učebny, atd. Zároveň ale může u 'highstake' testů nastat problém se zabezpečením.
- Rozdíly v zadání a instrukcích – instrukce jsou většinou u počítačových testů jednotné a konzistentní, ovšem Douglas a Chapelle poukazují na fakt, že tzv. okna s nápovědou (help screens) mohou být pro studenty matoucí či rušivé.
- Rozdílný obsah a formát odpovědí – Počítače umožňují více podob a možností obsahu, což vede k větší autentičnosti, ale závisí to na dostupných technologiích.
- Vztah mezi obsahem a formátem – Technologie nám nabízí speciální formu počítačového testování, tzn. 'Computer Adapted Testing', při které dochází k přesnějšímu měření, protože testovaný subjekt dostává otázky podle úspěšnosti předešlých odpovědí, tudíž přiměřené jeho znalostem.
- Rozdíly v hodnocení – automatizované hodnocení urychluje celý testovací proces, k dispozici jsou snadno i statistické údaje, položková analýza, atd. Nicméně hodnocení otázek s tvořenou odpovědí zůstává problematické a lidský hodnotitel je stále potřeba.

1.3.2 PROBLEMATICKÉ ASPEKTY VALIDITY U POČÍTAČOVÉHO TESTOVÁNÍ

- Chapelle a Douglas nastiňují různé problematické aspekty ohledně validity počítačového testování a snaží se i o reakce na tyto problémy.

Možný problematický aspekt validity	Vysvětlení	Reakce
<i>Rozdílný výkon</i>	Počítačové testy nemusí testovat stejné dovednosti jako testy papírové.	Pokud je tomu tak, je potřeba výzkumu zaměřeného na rozdíly výkonu u těchto různých forem testování.
<i>Nové typy úloh</i>	Typy úloh jsou u počítačových testů odlišné.	Kvalitativní a kvantitativní výzkum musí potvrdit vhodnost nových testovacích úloh.
<i>Omezení kvůli adaptivnímu módu testování (CAT)</i>	Vzorčky položek nemusí správně reflektovat obsah testovaného jevu.	Je potřeba prozkoumat různé formy adaptivního testování.
<i>Nepřesné automatizované hodnocení</i>	Automatizované hodnocení neobdohuje uznatelné odpovědi, které nejsou v klíči.	Přístrojové hodnocení musí být správně využíváno.
<i>Ohrožení zabezpečení</i>	Počítačové testování může přestavovat riziko pro zabezpečení.	Jednotlivá zabezpečení musí odpovídat účelu testu.
<i>Negativní následky</i>	Počítačové testování může mít negativní dopad na studenty, učení, třídu i společnost	Porozumění a plánování dokumentace k možným negativním důsledkům je nutností.

(Převzato a adaptováno z Chapelle & Douglase 41-61)

- Jak je patrné z tabulky výše, počítačové testování s sebou přináší mnoho problematických aspektů které by měly být detailněji prozkoumány.

1.3.3 EFEKTIVITA VYUŽITÍ POČÍTAČŮ V TESTOVÁNÍ

- Brown přehledně uvádí výhody a nevýhody počítačového testování.
- Výhody dále rozděluje do dvou oblastí – z pohledu studenta a z pohledu samotné testovací situace.

STUDENT	TESTOVACÍ SITUACE
Studenti mohou pracovat svým vlastním tempem.	Počítače jsou spolehlivější při reportování výsledků.
V počítačovém adaptivním testování (CAT) se studenti cítí méně frustrováni, jelikož pracují na položkách, které odpovídají jejich úrovni znalostí.	Počítače umožňují okamžitou zpětnou vazbu a statistická data, včetně diagnostického feedbacku.
Počítačové testování zabírá méně času a je efektivnější.	Počítačové adaptivní testy a IRT umožňují přesnější ohodnocení úrovně znalostí.
Počítačové testování je méně vyčerpávající, jelikož položky jsou představovány postupně, nikoli najednou.	Počítače jsou přesnější v hodnocení než lidé.
Studenti mají k moderním technologiím blízko a někteří si i proces testování na počítačích užijí.	Dostatečně velké banky úloh a různé testy by měly minimalizovat podvádění.

- Na nevýhody se Brown dívá z hlediska fyzického a výkonnostního.

FYZICKÉ hledisko	VÝKONOSTNÍ hledisko
Nedostatečná dostupnost počítačového vybavení	Různé výsledky závislé na způsobu administrace

Omezená kapacita obrazovky může být problematická	Rozdílná počítačová gramotnost znevýhodňuje studenty nedostatečně zběhlé v počítačovém testování
Možnosti grafiky mohou být omezené	Nervozita z počítačů může negativně ovlivnit výsledek

(Převzato a adaptováno z Browna 47,48)

- Z tabulek je patrné, že výhody jasně nad nevýhodami převažují.
- Brownův výčet pochází z roku 1997, a podle mého názoru se moderní technologie od té doby velmi vyvinuly, takže např. omezená grafika či kapacita obrazovky už by dnes neměly být problémem. Brownův seznam je v předkládané disertaci doplněn a kontrastován s mnoha poznatky dalších autorů publikovaných po roce 2000.