

# Lze žákům zeměpisu pomoci k lepším výkonům změnou řazení testových úloh? – případová studie na studentech Lauderových škol

VÁCLAV STACKE, KLÁRA VOČADLOVÁ, RADEK REICHERT

**G** **Abstrakt:** Vytváření testů je důležitým nástrojem pro hodnocení porozumění žáků a jejich kompetencí k učení. Existuje mnoho různých přístupů k testování a mnoho typů testových otázek, které mohou ovlivnit celkové hodnocení žáka. Může být však celkový výsledek ovlivněn i samotným pořadím testových úloh? Tento příspěvek se zabývá vlivem uspořádání otázek na výsledky testů ze zeměpisu u 97 žáků šesti tříd osmiletého gymnázia Lauderových škol v Praze. Vytvořeny byly tři varianty testů – chronologické pořadí, od nejjednodušší po nejtěžší a náhodně řazené otázky. Každý žák byl testován každou z variant. Dosažený počet bodů v jednotlivých variantách byl statisticky analyzován. Nebylo prokázáno, že by řazení otázek mělo signifikantní vliv na celkový počet bodů získaných žáky. Dříve než bude možné zhodnotit celkový význam uspořádání testových úloh na výkony žáků, bylo by třeba dalšího výzkumu zaměřeného zejména na propojení našich výsledků s výzkumy zaměřenými na oblast vlivu vnitřních determinantů (např. výkonová motivace, kognitivní styl žáka).

**Klíčová slova:** výuka zeměpisu, zeměpis, evaluace, testování, pořadí úloh, motivace.

STACKE, V., VOČADLOVÁ, K. & REICHERT, R. 2018. Lze žákům zeměpisu pomoci k lepším výkonům změnou řazení testových úloh? – případová studie na studentech Lauderových škol. *Arnica*, 8, 2, 69–75. Západočeská univerzita v Plzni, Plzeň, ISSN 1804–8366.

Rukopis došel 31. 1. 2018, přijat po recenzi 30. 11. 2018.

Václav Stacke, Centrum biologie, geověd a envigogiky, Fakulta pedagogická, Západočeská univerzita v Plzni, Klatovská 51, 306 19, Plzeň; email: stacke@cbg.zcu.cz • Klára Vočadlová, Centrum biologie, geověd a envigogiky, Fakulta pedagogická, Západočeská univerzita v Plzni, Klatovská 51, 306 19, Plzeň; email: vocadlo@cbg.zcu.cz • Radek Reichert, Mírové náměstí 1175/5, 405 02, Děčín; email: reichert.radek@gmail.com

## Úvod

Úspěšnost vzdělávacího procesu je ovlivněna mnoha jevy. Obecně lze za tyto jevy považovat vliv rodiny na žákovu připravenost, vliv učitele na vyučovací proces, vliv vnitřních a vnějších potřeb žáka, další motivaci atd. Dle klasické didaktické terminologie je hodnocení úspěšnosti vzdělávacího procesu sférou učitele. Učitel hodnotí vzdělávací proces žáka a vynáší hodnotící soudy – žák prospívá či neprospívá. Hodnocení je tak nedílnou součástí vzdělávacího procesu. Učitelova úloha hodnocením nekončí, na základě zjištěných výsledků učitel vzdělávací proces dále upravuje. Hodnocení je tak kontrolní fází toho, jak jsou naplňovány myšlené a následně realizované vzdělávací cíle.

V posledních desetiletích se problematice hodnocení vzdělávacího procesu v českém prostředí věnovala řada prací (např. Turek 1987, Slavík 1999, Skalková 1999, Kalhous, Obst *et al.* 2002, Průcha 2002, Kolář & Šikulová 2009, Mareš 2013, Zormanová 2014).

V případě vzdělávacího procesu a jeho hodnocení hraje významnou roli tzv. výkonová motivace jedince (Plháková 2004), která určuje, zda žák úlohu vyřeší, či se o její vyřešení vůbec pokusí. Moorhead & Griffin (1989, s. 119) dokládají, že motivace k dosažení úspěchu je ovlivněna zkušenostmi z minulosti. Motivace k participaci na všech etapách vzdělávacího procesu pak vychází z fundamentální hierarchie lidských potřeb (Maslow 1943), zejména

ze dvou nejvyšších potřeb, tedy potřeby uznání (*esteem*) a seberealizace (*self-actualization*). S motivací dále pracuje McClelland (1958) ve své teorii získaných potřeb, kde se ve vzdělávacím procesu a jeho hodnocení uplatňuje zejména *achievement motivation*, tedy potřeba něčeho dosáhnout, která však naráží na strach ze selhání. Atkinson (1964) popisuje dvě tendence – tendenci dosáhnout úspěchu a tendenci vyhnout se neúspěchu, které jsou určující pro aktivizaci žáka a jeho následný výkon. Heckhausen (1967) tyto myšlenky rozvedl do tvrzení, že pocit úspěchu roste, pokud pravděpodobnost úspěchu klesá, a pocit neúspěchu roste, pokud subjektivní obtížnost dané úlohy klesá. Nejistotu spatřuje v tom, že motivy nejsou přímo pozorovatelné a jejich exaktní diagnostické měření je problematické. Hrabal & Pavelková (2011) vytvořili standardizovaný dotazník, na základě jehož výsledků dělí žáky do pěti typů dle poměru mezi potřebou úspěšného výkonu a potřebou vyhnout se neúspěchu. Strach z neúspěchu tak některé skupiny žáků může vést až k apatii, která se projevuje neúspěchem ve vzdělávacím procesu, přestože má žák vnitřní potenciál uspět (Mareš 2013, s. 549).

Má-li žák pozitivní zkušenost s vlastním dosaženým výkonem, je u nové úlohy hnán touhou ji nejen splnit, ale zároveň ji splnit lépe. Pokud vyjdeme z předpokladu, že obtížnější úlohy vyžadují od žáka vyšší míru aktivizace

a úsilí, můžeme dovodit, že žák je vlivem kladné zkušenosti schopen zvládnout náročnější úlohu.

Formulace otázek v didaktickém testu může ovlivnit žákův výkon. Žák vnímá, jak jsou úlohy formulovány a jak náročné je jejich řešení. Každá úspěšně vyřešená úloha žáka motivuje k vyššímu nasazení při řešení další testové úlohy a tím dochází k nastartování spirály pozitivní zpětné vazby. Seřazení testových úloh dle náročnosti a stupňujících se nároků na schopnosti žáka tak může významně ovlivnit jeho motivaci při vypracování testu a tím i jeho výsledný výkon.

V zahraničí se s rozdílnými výsledky problematice řazení testových úloh a jejímu vlivu na motivaci v obecné rovině i vybraných předmětech věnovala celá řada prací, přehled starších z nich je podrobně zpracován (Leary & Dorans 1982), mezi novější studie pak patří např. Pettijohn & Sacco (2007), Weinstein & Roediger (2012). Studie se obecně zabývaly různými věkovými skupinami, testovaly vědomosti v různých oborech (psychologie, matematika, právo, geografie) a řadily otázky podle různých pravidel (podle obsahu, od nejjednodušší, od nejtěžší, náhodně). Z celkového pohledu nejsou však výsledky konzistentní, vliv řazení otázek na výkon studentů není ve všech studiích jednoznačně signifikantní. Na výkony žáků v testech mohou mít tedy vliv i další faktory jako je pohlaví testovaných osob, členění testů do tematických celků či způsob přípravy na test. Hambleton & Traub (1974) porovnávali výsledky žáků v testech z matematiky při řazení úloh od nejjednodušší po nejobtížnější a při obráceném řazení. Zjistili, že žáci dosahovali prokazatelně lepších výkonů při řazení od nejsnazších úloh po nejobtížnější. Weinstein a Roediger (2012) prokázali, že výkony žáků v testech obecných znalostí jsou při řazení úloh s vzrůstající obtížností vyšší, než když jsou úlohy řazeny od nejobtížnější po nejsnazší. Plake (1980) porovnával výkony žáků v testech, jejichž úlohy byly řazeny dle obtížnosti, cyklickou spirálou a náhodně, avšak průkazně nebylo možno určit, zda v jedné z variant žáci skutečně dosahují vyšších výkonů. Stejně i např. Pettitjohn & Sacco (2007) neprokázali statisticky signifikantní rozdíly ve skóre žáků v závislosti na typu testu. Toto ovšem neplatilo u vnímání obtížnosti jednotlivých testů samotnými žáky, kdy se prokázalo, že testy s otázkami řazenými náhodně byly označovány jako nejobtížnější.

V českém prostředí se problematikou řazení testových úloh zabývali Byčkovský (1982) a Chráska (1999), kteří axiomaticky hovoří o nutnosti řadit úlohy dle vzrůstající obtížnosti. Toto tvrzení se pak promítá i v didaktických vysokoškolských skriptech (Škoda *et al.* 2006, Kratochvílová 2015, Škaloudová 2015), ve kterých dochází k cyklické citaci. Studie prakticky ověřující vliv řazení testových

otázek na výkony žáků v zeměpisu však v českém prostředí prozatím zcela chybí.

Tento text si za cíl klade evaluovat rozdíly ve výkonech žáků v didaktických testech ze zeměpisu při rozdílném řazení shodně formulovaných otázek. Výzkumnou hypotézou je, že studenti dosahují vyšších výkonů v didaktickém testu, pokud má test úlohy řazeny dle vzrůstající obtížnosti, než při řazení chronologickém či náhodném.

Jako v každém jiném hodnocení výkonu ve vzdělávacím procesu, je nedílnou součástí i této studie kromě objektu – žáka, i subjekt – učitel. Od učitele je očekáván objektivní a nestranný přístup v hodnocení vzdělávacího procesu, k čemuž mu pomáhají didaktické nástroje. Tento text přispívá k lepšímu pochopení možností jednoho z tradičních didaktických nástrojů, a to didaktického testu.

## Metodika

Pro zjištění výkonu žáků v didaktických testech při různém řazení shodně formulovaných testových úloh jsme postupovali experimentální strategií za použití empirických dat kvantitativního rozsahu. K zajištění objektivních kritérií sběru dat jsme použili jasně definovaný didaktický nástroj – didaktický test. Metodologie výzkumné části byla zpracována dle Chrásky (1991), Švece *et al.* (2009) a Gavory (2010). Použité matematicko-statistické metody vychází z prací Procházky (1993), Hrachy (2006) a Budíkové a kol. (2010).

Každý z didaktických testů předložených žákům byl sestaven na základě realizovaného kurikula v konkrétní třídě. Test následoval vždy po 3–4 týdnech výuky a byl žákům ohlášen týden předem.

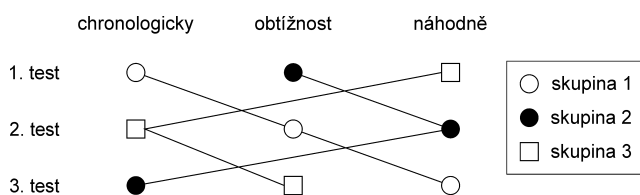
Pro tvorbu každého testu byla dle náročnosti tématu vždy vypracována databáze 20–30 testových úloh, a následně z ní bylo vybráno 10 úloh, které byly použity v testu. Úlohy byly vybrány tak, aby rovnoměrně pokryly celou šíři testovaného tématu, a současně aby byly vyváženě zastoupeny jednotlivé typy úloh. Obtížnost testových úloh byla volena dle schopností žáků v daných třídách a úrovně vzdělávacích cílů daného ročníku. Každá testová úloha byla hodnocena 10 body, maximálně tak žáci mohli dosáhnout v každém testu 100 bodů. Při projektování každého z testů byl brán zřetel na jeho objektivitu, validitu, reliabilitu i praktičnost. Vzniklý test byl následně upraven do tří variant dle způsobu řazení otázek.

Ve variantě A (chronologické řazení) byly úlohy seřazeny chronologicky, kdy úlohy, testující nejdříve probírané učivo, byly zařazeny v úvodu testu. Pro variantu B (řazení dle obtížnosti) byly úlohy řazeny dle vzrůstající obtížnosti od nejsnazší. Přestože určení obtížnosti je vždy do jisté míry subjektivní, byly u každé z úloh pro co nejvyšší

objektivitu brány v úvahu složitost větné vazby, použití cizích slov a nápověd v zadání a náročnost vzdělávacího cíle (sensu Bloom et al. 1956). Ve variantě C (náhodné řazení) byly otázky seřazeny náhodně, kdy každé z otázek bylo v programu MS Excel matematickou funkcí NÁHČÍSLO přiřazeno náhodně generované číslo a následně byly otázky seřazeny vzestupně podle hodnoty tohoto čísla. Testy byly sestaveny tak, aby jejich index obtížnosti P (dle Chráska 1991) nabýval hodnoty  $<20; 80>$ .

Žáci v každé třídě byli rozděleni na tři skupiny pomocí matematické funkce NÁHČÍSLO v programu MS Excel. Každému žákovi bylo přiřazeno náhodné číslo, následně byli dle hodnoty tohoto čísla žáci seřazeni a posloupenosti 1-2-3-1-2-3 atd. rozřazeni do tří skupin. Každý z žáků se v různém pořadí setkal s každou z variant řazení testových úloh. Řazení jednotlivých variant testů pro každou ze skupin je zobrazeno na obr. 1.

Zapojení žáci byli anonymizováni přiřazením alfanumerického kódu, který určuje konkrétního žáka, jeho zařazení do konkrétní třídy a skupiny řazení testových variant. Tato anonymizace proběhla z důvodu snížení možné subjektivity při opravování testů, kdy učitel nedokázal snadno přiřadit test konkrétnímu žákovi, a zároveň i z etických důvodů – při dalším zpracování výsledků již není známa identita konkrétních žáků.



**Obr. 1.** Systém řazení jednotlivých variant testů u jednotlivých skupin žáků. Je patrné, že každý žák se setkal s každou z variant právě jednou.

Data získaná z výsledků testů byla v programu MS Excel uspořádána do frekvenční tabulky. Hodnoty sledovaných proměnných byly převedeny do frekvenční tabulky, která byla vybrána jako adekvátní nástroj k selekci velkého množství dat dle předem definovaných vzorců a provedení vybraných početních úkonů. Takto byly získány průměrné hodnoty pro výkony žáků v každé třídě, ve vybraném testu a v dané skupině. Celkové výkony žáků pak byly statisticky zpracovány v programu STATISTICA, kde byla výchozí data testována Kruskal-Wallisovým testem (KW-H). Výsledky byly znázorněny graficky pomocí krabicového grafu (Obr. 2).

Nulová hypotéza zněla: *Nebudou existovat rozdíly v rozptylu hodnot uvnitř souborů dat, tzn. řazení otázek nemá vliv na výkony žáků.* Testování bylo provedeno na základě analýzy vztahů mezi rozptyly v jednotlivých skupinách.

## Charakteristika výzkumného vzorku

Výzkum byl prováděn na Lauderově MŠ, ZŠ a gymnáziu při Židovské obci v Praze. Výuka na této instituci je přizpůsobena specifikům žáků ze židovských rodin, školní vzdělávací program (ŠVP) Le Chajim přesto naplňuje vzdělávací cíle rámcových vzdělávacích programů (RVP ZŠ a RVP G) MŠMT pro všechny tři vzdělávací stupně. Část učiva, tradičně vyučovaného v zeměpisu, je v kvartě (devátém ročníku) dle ŠVP Lauderových škol integrována do samostatného předmětu Svět v souvislostech.

Počty žáků, struktura třídy a týdenní hodinové dotace zeměpisu v jednotlivých ročnících, ve kterých byl výzkum prováděn, jsou uvedeny v tab. 1. Žáci septimy, ve které je dvouhodinová týdenní dotace, nebyli do výzkumu zařazeni.

Ročník	Týdenní hodinová dotace zeměpisu	Celkový počet žáků	Počet chlapců	Počet dívek
Prima	2	17	8	9
Sekunda	1	18	9	9
Tercie	1	20	10	10
Kvarta	1	15	1	14
Kvinta	2	18	7	11
Sexta	2	9	4	5

**Tab. 1.** Počty žáků a hodinová dotace zeměpisu v jednotlivých ročnících, ve kterých byl výzkum prováděn.

Z nestrukturovaných rozhovorů s třídními učiteli nevyplývala žádná specifika, která by se vymykala běžnému psychosociálnímu vývoji žáků.

## Výsledky a diskuse

U všech testů v jednotlivých ročnících byl na základě výkonů žáků spočten jejich index obtížnosti P (sensu Chráska 1991), který je uveden v tab. 2. Indexy všech testů se nacházejí v intervalu  $<20; 80>$ , takže byly sestaveny s vhodnou obtížností. Krajiním hodnotám se blíží pouze první test v sextě a druhý test v primě, kdy oba testy byly na samé hranici, za kterou by již byly příliš snadné. V tab. 3 jsou pro každý ročník uvedeny počty testů každé varianty, ve kterých testování žáci dosáhli nejvyšších průměrných výkonů. Tab. 4 pak zobrazuje základní statistické ukazatele výkonů žáků v jednotlivých variantách testů s rozdílným řazením testových úloh.

Rozdíly v průměrném hrubém skóre žáků pro jednotlivé varianty testů jsou vyobrazeny na obr. 2.

	Prima	Sekunda	Tercie	Kvarta	Kvinta	Sexta
Test č. 1	67,1	66,2	66,6	53,3	41,4	80
Test č. 2	79,7	64	64,5	57,6	59,3	64,4
Test č. 3	59,9	69,2	46,4	62,1	71,1	75,7

**Tab. 2.** Index obtížnosti P (sensu Chráska 1991) jednotlivých testů (bez rozlišení varianty řazení testových úloh) v jednotlivých ročnících.

	Chronologicky	Dle obtížnosti	Náhodně
Prima	2	1	0
Sekunda	1	1	1
Tercie	1	1	1
Kvarta	1	2	0
Kvinta	0	1	2
Sexta	1	2	0

**Tab. 3.** Úspěšnost skupin žáků jednotlivých ročníků podle testových variant. Číslo ukazuje, kolikrát skupina žáků vypracovávající danou variantu dosahovala nejlepších výsledků ze všech variant testů. Např. v primě ze tří termínů testů měli dvakrát nejlepší průměrný výkon žáci, kteří zrovna vypracovávali variantu testu s úlohami seřazenými chronologicky, a jednou žáci s variantou testu s úlohami řazenými dle obtížnosti. Žáci s variantou testu s úlohami řazenými náhodně nedosáhli nejlepšího průměrného výkonu ani jednou.

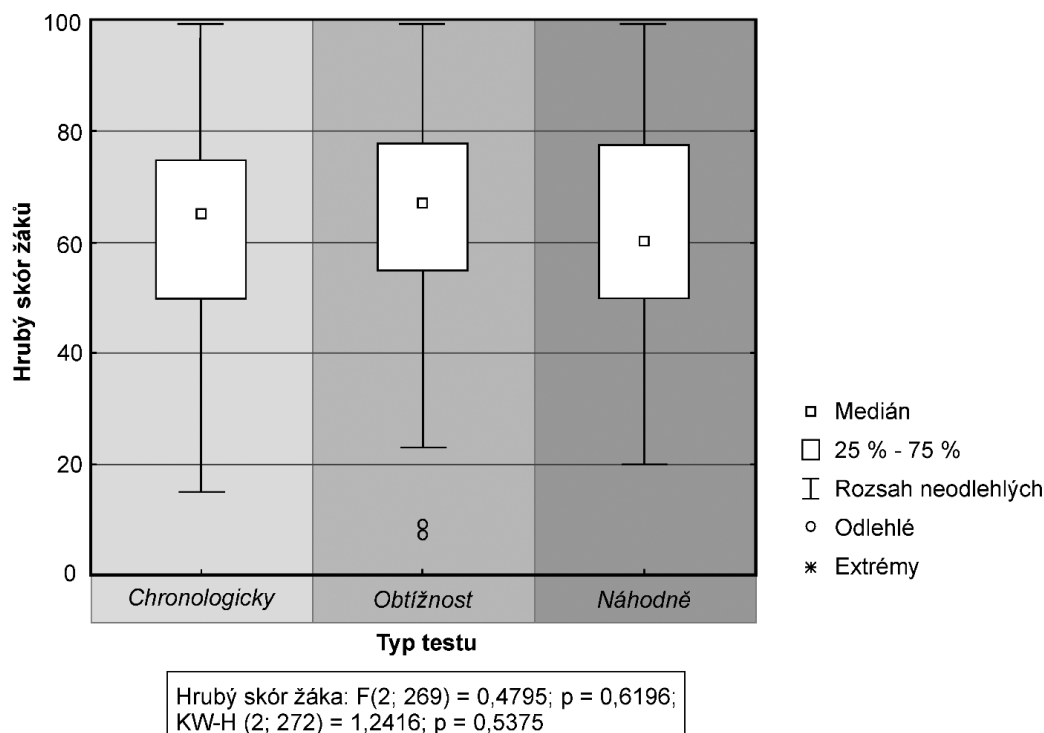
Při řazení dle vzrůstající obtížnosti celý testovaný vzorek žáků dosáhl průměrného výkonu o 2,3 % vyššího, nežli při řazení chronologickém, a o 3,2 % vyššího, nežli při řazení náhodném. Podobných výsledků dosahují i hodnoty mediánu – v porovnání chronologického řazení a řazení dle obtížnosti dosahuje rozdíl 4 %, při porovnání s náhodným řazením dokonce 6 %.

Přestože tyto výsledky vypadají atraktivně a měly by odpovídat zjištěním Hambletona & Trauba (1974) či Weinsteina & Roedigera (2012), že pokud jsou úlohy v testu řazené se vzrůstající náročností, dosahují žáci v této variantě vyšších výkonů, na základě hlubší statistické analýzy musíme podpořit spíše studie, které poukázaly na statistickou neprůkaznost změn výkonů žáků v testech při rozdílném řazení úloh (např. Plake 1980, Pettijohn & Sacco, 2007). Rozdíl mezi rozptyly testovaných souborů je dle našich analýz statisticky nevýznamný ( $p > 0,05$ ), variabilita výkonů žáků v různých typech testů se tedy významně neliší. Ani Kruskal-Wallisův test neprokázal vliv změny řazení testových otázek na výkon žáků.

		Chronologicky	Dle obtížnosti	Náhodně
Prima	Minimum	52,3	45	47,3
	Medián	<b>72,2</b>	68,3	64,3
	Maximum	91,7	<b>94,3</b>	89,3
	Průměr	<b>72,1</b>	69,2	65,9
Sekunda	Minimum	44,7	42	46
	Medián	65,8	<b>74,8</b>	68,2
	Maximum	84,3	83	<b>85,7</b>
	Průměr	65	67,2	<b>67,3</b>
Tercie	Minimum	42,3	26	40
	Medián	57,3	<b>60,5</b>	54,7
	Maximum	85,3	<b>87,3</b>	82,3
	Průměr	<b>59,4</b>	59,3	59
Kvarta	Minimum	30	46,5	24
	Medián	59	<b>62,5</b>	54
	Maximum	<b>75</b>	72,5	73
	Průměr	56,7	<b>60</b>	49,9
Kvinta	Minimum	40	33,3	36
	Medián	54,3	<b>62,3</b>	54,2
	Maximum	73,3	<b>86,3</b>	78,7
	Průměr	55,4	<b>60,7</b>	56,1
Sexta	Minimum	56	56,3	67,3
	Medián	74,7	<b>79</b>	76
	Maximum	81	<b>90</b>	80
	Průměr	70,6	<b>75,1</b>	74,4
Celék	Minimum	<b>44,2</b>	41,5	43,4
	Medián	63,9	<b>67,9</b>	61,9
	Maximum	81,8	<b>85,6</b>	81,5
	Průměr	63,2	<b>65,3</b>	62,1

**Tab. 4.** Statistické ukazatele výkonu žáků v jednotlivých ročnících a celého testovaného vzorku dle variant testů s rozdílným řazením testových úloh. Nejvyšší hodnoty jsou zvýrazněny.





**Obr. 2:** Krabicový diagram (Box plot) znázorňuje rozložení dosaženého celkového hrubého skóre u všech žáků ve všech testech dle tří typů řazení testových úloh (chronologicky, dle obtížnosti, náhodně). Linie vycházející ze střední části boxu představují minimální a maximální hodnoty hrubého skóre. Spodní a svrchní část boxu ukazují 1. a 3. kvartil, značka uprostřed boxu vyznačuje medián.

Z uvedeného statistického zhodnocení vyplývá, že rozdíly v dosažených výsledcích v závislosti na řazení otázek v testu nejsou statisticky významné (F-test, obr. 2).

I přesto však můžeme přijmout tvrzení, že řazení testových úloh může napomoci snížit hladinu stresu během plnění testových úloh a mírně zvyšovat motivaci, jak uvádějí Byčkovský (1982) či Chráska (1999).

Při pohledu na výsledky jednotlivých testovaných tříd je situace ještě komplikovanější. Např. v primě je možné vysledovat, že žáci dosahovali lehce lepších nebo shodných výsledků při chronologickém řazení otázek. Taktéž je třeba brát v úvahu, že aktuální výkon žáka mohou ovlivňovat různé faktory – ať už klima ve škole a ve třídě, rodinné prostředí, momentální psychické rozpoložení či zdravotní stav (Vališová & Kasíková 2007). Velmi důležitými faktory, ovlivňujícími žákovy výkony jsou pozornost a délka soustředění, které jsou žáci schopni dosáhnout (Petty 1996, Mareš 2013, Vágnerová 2014). Otázkou tak i nadále zůstává velikost vlivu jednotlivých faktorů na žákovy výkony v didaktických testech.

## ■ Závěr

Porovnáním výkonů 97 žáků šesti tříd Lauderových škol v Praze se nám nepodařilo statisticky průkazně prokázat, že pokud jsou úlohy v didaktickém testu ze zeměpisu seřazeny se vzrůstající obtížností, dosahují

žáci vyšších výkonů, nežli při řazení úloh náhodně, či chronologicky. Přestože průměrné hodnoty výkonů za celý testovaný vzorek byly při řazení testových úloh dle obtížnosti v řádech jednotek procent vyšší, než při řazení chronologicky či náhodně, signifikantní závislost se nepotvrdila. V ojedinělých případech byly výkony jednotlivců či celých ročníků vyšší v testech s chronologicky řazenými úlohami. Pro verifikaci či definitivní vyvrácení hypotézy je tak třeba dalšího výzkumu. Kromě výzkumu většího vzorku tříd z různých typů škol napříč celým vzdělávacím systémem, je pak nutné zejména propojení našich výsledků s výzkumy zaměřenými na oblast vlivu vnitřních determinantů, jako je např. výkonová motivace a kognitivní styl žáka.

## ■ Poděkování

Text vznikl za finanční podpory projektu OPVVV Didaktika – Člověk a příroda A (CZ.02.3.68/0.0/0.0/16\_011/0000665), který je spolufinancován Evropskými strukturálními a investičními fondy a státním rozpočtem České republiky.

## Literatura

- ATKINSON, J. W. 1964. *An Introduction to Motivation*. Van Nostrand, Oxford. 335 pp.
- BLOOM, B. S., ENGELHART, M. D., FURST, E. J., HILL, W. H. & KRATHWOHL, D. R. 1956. *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. David McKay Company, New York. 207 pp.
- BUDÍKOVÁ, M., KRÁLOVÁ, M. & MAROŠ, B. 2010. *Průvodce základními statistickými metodami*. Grada, Praha. 272 pp.
- BYČKOVSKÝ, P. 1982. *Základy měření výsledků výuky: Tvorba didaktického testu*. ČVUT, Praha. 149 pp.
- GAVORA, P. 2010. *Úvod do pedagogického výzkumu*. Paido, Brno. 260 pp.
- HAMBELTON, R. K. & TRAUB, R. E. 1974. The effects of item order on test performance and stress. *The Journal of Experimental Education* 43 (1): 40–46.
- HECKHAUSEN, H. 1967. *The Anatomy of achievement motivation*. Academic Press, New York. 236 pp.
- HRABAL, V. & PAVELKOVÁ, I. 2011. *Školní výkonová motivace žáků. Dotazník pro žáky*. Národní ústav odborného vzdělávání, Praha. 27 pp.
- HRACH, K. 2006. *Sbírka úloh ze statistiky*. FSE UJEP, Ústí nad Labem. 65 pp.
- CHRÁSKA, M. 1991. *Metodologie řešení vybraných problémů v pedagogickém výzkumu*. Univerzita Palackého v Olomouci, Olomouc. 70 pp.
- CHRÁSKA, M. 1999. *Didaktické testy. Příručka pro učitele a studenty učitelství*. Paido, Brno. 91 pp.
- KALHOUS, Z., OBST, O. et al. 2002. *Školní didaktika*. Portál, Praha. 447 pp.
- KOLÁŘ, Z. & ŠIKULOVÁ, R. 2009. *Hodnocení žáků*. Grada, Praha. 200 pp.
- KRATOCHVÍLOVÁ, L. 2015. *Didaktický test – hodnocení*. [online] Pedagogická fakulta, Masarykova univerzita [cit. 14. 12. 2017]. Dostupný na WWW: <[http://is.muni.cz/el/1441/podzim2005/ZS1BK\\_PDD/XII.\\_Didakticky\\_test.doc](http://is.muni.cz/el/1441/podzim2005/ZS1BK_PDD/XII._Didakticky_test.doc)>.
- LEARY, L. F. & DORANS, N. J. 1982. The effect of item rearrangement on test performance: a review of the literature. *ETS Research Report Series 1982* (2): i–23.
- MAREŠ, J. 2013. *Pedagogická psychologie*. Portál, Praha. 702 pp.
- MASLOW, A. H. 1943. A theory of human motivation. *Psychological Review* 50 (4): 370–396.
- MCCLELLAND, D. C. 1958. *Methods of Measuring Human Motivation*. In ATKINSON, J. W. (ed.) *Motives in Fantasy, Action and Society*. Van Nostrand, Princeton: 12–13.
- MOORHEAD, G. & GRIFFIN, R. W. 1989. *Organizational Behavior*. Houghton Mifflin, Boston. 735 pp.
- PETTITJOHN, T. F. & SACCO, M. F. 2007. Multiple-Choice Exam Question Order Influences on Student Performance, Completion Time, and Perceptions. *Journal of Instructional Psychology* 34 (3): 142–149.
- PETTY, G. 1996. *Moderní vyučování*. Portál, Praha. 380 pp.
- PLAKE, B. S. 1980. Item arrangement and knowledge of arrangement on test scores. *Journal of Experimental Education* 49 (1): 56–58.
- PLHÁKOVÁ, A. 2004. *Učebnice obecné psychologie*. Academia, Praha. 472 pp.
- PROCHÁZKA, V. 1993. *Úvod do matematické statistiky*. Západočeská univerzita v Plzni, Plzeň. 111 pp.
- PRŮCHA, J. 2002. *Moderní pedagogika*. Portál, Praha. 481 pp.
- SLAVÍK, J. 1999. *Hodnocení v současné škole*. Portál, Praha. 190 pp.
- SKALKOVÁ, J. 1999. *Obecná didaktika*. Nakladatelství ISV, Praha. 292 pp.
- ŠKALOUDOVÁ, A. 2015. *Ověřování testu – položková analýza*. [online] Pedagogická fakulta, Univerzita Karlova v Praze [cit. 14. 12. 2017]. Dostupné na WWW: <<http://kps.pedf.cuni.cz/skalouda/Psychometrie/psychometrie.htm>>.
- ŠKODA, J., DOULÍK, P. & HAJEROVÁ-MÜLLEROVÁ, L. 2006. Zásady správné tvorby, použití a hodnocení didaktických testů v přípravě budoucích učitelů. [online] ŠKODA, J., DOULÍK, P. & HAJEROVÁ-MÜLLEROVÁ, L. *Interaktivní cvičebnice tvorby a hodnocení didaktických testů*. [cit. 14. 12. 2017] Dostupné na WWW: <<http://cvicebnice.ujep.cz/cvicebnice/FRVS1973F5d/>>.
- ŠVEC, Š. et al. 2009. *Metodologie věd o výchově. Kvantitativně-scientické a kvalitativně-humanitní přístupy v edukačním výzkumu*. Paido, Brno. 302 s.
- TUREK, I. 1987. *Didaktika technických predmetov*. Slovenské pedagogické nakladateľstvo, Bratislava. 178 pp.
- VÁGNEROVÁ, M. 2014. *Vývojová psychologie. Dětství a dospívání*. Karolinum, Praha. 536 pp.
- VALIŠOVÁ, A. & KASÍKOVÁ, H. 2007. *Pedagogika pro učitele*. Grada, Praha. 402 pp.
- WEINSTEIN, Y. & ROEDIGER, H. L. 2012. The effect of question order on evaluations of test performance: How does the bias evolve? *Memory & Cognition* 40 (5): 727–735.
- ZORMANOVÁ, L. 2014. *Obecná didaktika*. Grada, Praha. 240 pp.

## E English summary

### Can we help students of geography to achieve better results by changing test tasks order? – case study of 97 students from Lauder school

Designing test is an important tool for assessing pupils understanding and evaluating of their learning competencies. There are many different testing methods and different types of examination questions. These could influence the final score. Could the final score depend on the questions order? This contribution evaluates the effect of question rearrangement on geography test performance. The research sample consisted of six classes at an eight-year secondary school in Prague. Three question arrangements existed – chronological order, easy-to-hard, and randomly arranged questions. Each pupil was tested using all three item arrangements. The overall gross score that students achieved in the tests was analysed. There was no statistically significant effect of item rearrangement on the score. We cannot definitely confirm the positive effect of item arrangement on pupil performance. The further research focused especially on the influence of internal determinants (such as achievement motivation and students cognitive style) is needed before a conclusion of item arrangement significance can be made.

**Key words:** geography education, geography, evaluation, testing, tasks order, motivation.

#### Figures

**Fig. 1.** Arranging of different test designs in particular group of students. Each student has passed all three variants of test once.

**Fig. 2.** Box plot represents general gross score that pupils achieved in all test types (chronologically, increasing difficulty, randomly arranged). Whiskers indicate the minimum and maximum of gross score. The bottom and top of the box show the first and third quartiles, and the mark inside the box is the median.

#### Tables

**Tab. 1:** Teaching time of geography per week (taught time in periods of 45 minutes) and number of pupils per class (total, boys, girls).

**Tab. 2:** Item difficulty index P (sensu Chráska 1991) for all tests and all classes (test type is not indicated).

**Tab. 3:** Success rate of pupils according to the test arrangement (chronological, increasing difficulty and random arrangement). Value indicates how many times the specific arrangement reported the best results in test scores. E. g., the group of first year pupils (prima) with chronological item arrangement achieved the higher score than pupils with other test arrangements twice, the first year pupils with easy-to-hard arrangement once and the pupils with random arrangement not once.

**Tab. 4:** Statistical variables of pupil performance (minimum, median, maximum, mean) according to the test arrangement. The highest values are highlighted.