

**Posudek práce předložené k obhajobě
na Matematicko-fyzikální fakultě Univerzity Karlovy**

- posudek vedoucí/vedoucího posudek oponentky/oponenta
 bakalářské práce diplomové práce

Název práce: Psychometrická analýza Lawsonova testu vědeckého uvažování

Jméno, příjmení a tituly autorky/autora práce: Bc. Martin Landa

Studijní program: Učitelství fyziky pro střední školy

Rok odevzdání: 2023

Jméno a tituly vedoucí/ho nebo oponentky/-ta: RNDr. Alexandr Nikitin

Pracoviště: Katedra didaktiky fyziky

E-mail: alexandr.nikitin@matfyz.cuni.cz

Hodnocená oblast	Hodnocení
Oborová úroveň (zejména fyzikální)	Velmi dobrá
Didaktická úroveň (zejména metody sběru a analýzy dat v didaktickém výzkumu, přiměřenost vzniklých materiálů)	Velmi dobrá
Práce s literaturou	Velmi dobrá
Jazyková úroveň práce (srozumitelnost textu, členění textu, stylistika, pravopis)	Průměrná
Grafická úroveň práce (formátování textu, typografie, přehlednost tabulek, kvalita obrázků, videí)	Průměrná
Zásady pro vypracování práce	Splněny

Slovní hodnocení (včetně upřesnění případných výhrad; mj. lze posoudit originalitu a kreativitu zvolených řešení, náročnost tématu apod.):

Předložená práce je zajímavým příspěvkem k fungování jednoho z nejnámějších testů vědeckého myšlení. Koncepte a rozsah práce odpovídají standardu v oboru, logická linie textu je dobře zřetelná a text je vystavěn koherentně. Náročnost zvolené metodologie je nadstandardní a výhrady, které zde dále uvádím jsou ve své podstatě vesměs formálního charakteru. Po stránce odvedené práce, a za ní ležícího výzkumného procesu považuji předloženou diplomovou práci za kvalitní, jsou ovšem náležitosti, které by ji jakožto vědeckou práci mohly vydatně pozvednout.

Jedinou věcnou chybou, na kterou jsem narazil, je popis *Baseline-category Logit Nominal response modelu*, kde se tvrdí, že využívá stejné dva parametry jako 2PL model. Tento model sice využívá tyto dva parametry, ale pro každou hodnotu nominální odpovědi využívá vlastní dvojici (tedy jich celkem používá výrazně více), navíc jejich význam je oproti 2PL modelu trochu odlišný, závisí na tom, která odpověď je zvolena jako Baseline (to, jak je s ním pracováno v této práci to ovšem nijak neovlivňuje).

Velice oceňuji, že autor v práci pouze nepředkládá hrubá data, ale pouští se do kvalitní distraktorové analýzy, a hledá příčiny a miskoncepce, které mohou být schovány za volbou konkrétních odpovědí. Porovnání výsledků s jinými autory je velice precizní, a přitom příjemně stručné. Závěr je formulován jasně a srozumitelně, jsou zde dokonce obsažena doporučení pro praxi.

Práce je napsána v odborném jazykovém stylu, který ovšem místy pokulhává. Kupříkladu v popisech obrázků je místo standardního termínu *charakteristická křivka položky* volen velice hrubý překlad "*grafy křivek předmětové charakteristiky*", přičemž třeba název celé *Teorie odpovědi na podložku* je zde ovšem užít pouze v angličtině (*Item response theory*). Místy jsou slova dvakrát, místy pro změnu chybí, dobrým příkladem je v předposledním odstavci na str. 19 celá věta "Z grafu vidíme...", která zde zjevně nepatří (a patří na str. 25, kde se také nachází).

Sic práce obsahuje velké množství tabulek a grafů, na naprostou většinu z nich není v textu explicitě odkazováno. Dále jsou tabulky vloženy jako obrázky a u grafů s mnoha křivkami by bylo vhodné volit jednoznačnější barevnou paletu.

Souhrnem bych chtěl konstatovat, že autor předloženou práci prokázal svou odbornost v tématu, a stejně tak své výzkumné kompetence.

Detailnější výhrady k jednotlivým částem práce (není nutné číst při obhajobě):

Má-li autor tyto údaje k dispozici, bylo by na místě na začátku druhé kapitoly uvést trochu detailnější popis zkoumaného vzorku žáků.

Tak jak je v kapitole 2.2 definován ULI index, by byl vždy nula (není totiž zřejmé, že ono UG a LG jsou počty **správných** řešení v daných skupinách respondentů). Dále je zde špatně označen rozptyl celého testu; diskriminace a obtížnost v 2PL modelu jsou popsány pomocí vlastností charakteristické křivky, ale nikde není popsáno, co je to charakteristická křivka; není zde popsán význam těchto parametrů.

Na začátku třetí kapitoly postrádám souhrnné výsledky testu, či přehled stručné položkové analýzy. Kdyby totiž například některé položky byly adepty na vypuštění (např. na základě změny Cronbachova alfa, kterou autor stejně určuje), je třeba je vyloučit předtím, než jsou provedeny IRT analýzy (které provádějí výpočty na všech zahrnutých položkách). Popřípadě, kdyby souhrnné skóre v celém testovaném souboru bylo razantně příkloněno k jednomu z extrémů, je význam indexů ULI či korelace položky s celkovým skóre přinejlepším diskutabilní.

Chtěl bych zde vyzdvihnout způsob, jakým jsou informace obsažené v charakteristických křivkách BLIRT modelu využity při interpretaci distraktorů, a to včetně souvislostí mezi úlohami. Neustálé opakování prakticky téhož grafu logistické funkce (byť s příslušnými parametry odhadnutého 2PL modelu) považuji za zbytečné, zajímavé by místo toho mohlo být porovnání těchto parametrů mezi položkami, např. v podobě tabulky (třeba na onom začátku třetí kapitoly). Místa jsou zde voleny nevhodné slovní formulace, které naznačují něco jiného, než autor zamýšlel – např. větu "S ohledem na to a na velmi slabé negativní korelace v případě zvolení distraktorů má schopnost resp. neschopnost vyřešení této úlohy relativně malý vliv na kognitivní vývoj řešitele," na straně 49 považuji za vyloženě zavádějící, proces kognitivního vývoje zde vůbec není předmětem zkoumání.

Při formulaci závěrů bych byl o něco opatrnější. Autor navrhuje zaměřit výuku zejména na kategorie myšlení, které silně korelují s celkovým skóre v testu. V tom vidím dva problémy – ponejprv korelace nezaručuje kauzalitu, nicméně autor jakožto možné rozšíření práce sám navrhuje ověření, zda by tyto změny vedly ke kýženému cíli. Za druhé – způsob jakým Lawsonův test operacionalizuje koncept "vědeckého myšlení" je jen jedním z mnoha způsobů, jak se to dá udělat, takováto rada by tedy mohla vést k vyučování typu "učíme se, abychom dobře dopadli v tomto konkrétním testu".

Případné otázky k obhajobě:

Jedním z předpokladů užitých IRT modelů je, že všechny položky měří jednu jedinou schopnost respondenta (unidimensionalita). Dokázal byste (například s přihlédnutím k tomu, že kategorie *stadium formálních operací* je zde charakterizována "schopností abstraktního myšlení, metodického postupu, vytváření hypotéz apod.") na příkladech konkrétních položek formulovat, jaké jiné schopnosti/dovednosti musí testovaný ovládat, aby v tomto testu dosáhl dobrého výsledku?

Velice oceňuji diskusi k úlohám 7 a 8. Distraktor *aa* je zde také docela významný, leč je při diskusi jen lehce zmíněn. Dokázal byste zanalyzovat proč je tato odpověď častá?

Předloženou práci

doporučuji

nedoporučuji

uznat jako práci bakalářskou/diplomovou.

Předloženou práci hodnotím stupněm: **Velmi dobře**

Datum a místo: V Praze dne 07.08.2023

Podpis: