

## METODIKA VYHODNOCENÍ VÝSLEDKŮ OVĚŘOVACÍHO TESTOVÁNÍ V POČÁTEČNÍM VZDĚLÁVÁNÍ

**Poskytovatel podpory:** Technologická agentura České republiky

Program na podporu aplikovaného společenskovedního a humanitního výzkumu, experimentálního vývoje a inovací ÉTA

**Projekt:** TL01000385 Metodika vyhodnocení výsledků ověřovacího testování v počátečním vzdělávání a její aplikace v modelových případových studiích

**Hlavní příjemce:** NEWTON College, a.s.

**Řešitel:** doc. RNDr. PhDr. Oldřich Hájek, Ph.D., MBA

**Autorský tým:** Mgr. Jiří Novosák, Ph.D., MBA; Ing. Jana Novosáková, Ph.D., MBA; Ing. Milan Lindner, Ph.D.; Ing. Blanka Vytrhlíková, MBA

ČR- Česká školní inspekce, Fráni Šrámka 37, 150 21 Praha 5

vydává

## OSVĚDČENÍ

č. j. ČŠIG-1253/21-G2, v návaznosti na interní předpis úřadu č. j. ČŠIG-3300/14-G21 a s odkazem na znění zákona č. 561/2004 Sb., o předškolním, základním, středním, vyšším odborném a jiném vzdělávání (školský zákon), ve znění pozdějších předpisů,

***o certifikaci dokumentu s názvem***

***Metodika vyhodnocení výsledků ověřovacího testování v počátečním vzdělávání***

**Řešitel:**

doc. RNDr. PhDr. Oldřich Hájek, Ph.D., MBA

**Autorský tým:**

Mgr. Jiří Novosák, Ph.D., MBA; Ing. Jana Novosáková, Ph.D., MBA; Ing. Milan Lindner, Ph.D.;  
Ing. Blanka Vytrhlíková, MBA

**Příjemce podpory, na jejímž základě byla metodika vytvořena:**

Newton College

**Projekt:**

TL01000385 Metodika vyhodnocení výsledků ověřovacího testování v počátečním vzdělávání a její aplikace v modelových případových studiích

**Popis:**

Metodika poskytuje uživatelům komplexní podpůrný nástroj s návodnými metodickými postupy vyhodnocení ověřovacích testů v počátečním vzdělávání. Metodika je určena organizacím, které se ve své činnosti zabývají vyhodnocováním ověřovacích testů v počátečním vzdělávání. Metodika může být využita při vyhodnocování ověřovacích testů v různých kontextových situacích a pro zpracování výstupů na různých úrovních hodnocení. Metodika zároveň umožňuje hodnotit kvalitu samotných ověřovacích testů (např. kvalita testových položek, soulad pozorovaných a modely predikovaných dat) a tím i robustnost formulovaných zjištění.

V Praze dne 9. března 2021

PhDr. Ondřej Andrys, MAE, MBA, MPA

náměstek ústředního školního inspektora

**ČSI**

Česká školní  
inspekce

Fráni Šrámka 37 | 150 21 Praha 5  
IČ: 00638994



## Oponenti

prof. PhDr. Jaroslav Veteška, Ph.D., MBA

doc. RNDr. Aleš Ruda, Ph.D.

PhDr. Eva Jurášková, Ph.D., MBA

**Prohlášení o poskytovateli dotace na projekt a projektu, v rámci něhož metodika vznikla**

Metodika byla zpracována jako výstup řešení projektu Technologické agentury České republiky číslo TL01000385 s názvem „Metodika vyhodnocení výsledků ověřovacího testování v počátečním vzdělávání a její aplikace v modelových případových studiích“, a to v rámci programu TL – Program na podporu aplikovaného společenskovedního a humanitního výzkumu, experimentálního vývoje a inovací ÉTA. Řešitelé projektu děkují Technologické agentuře České republiky za finanční podporu při řešení projektu.

## Obsah

1. Úvod.....	4
1.1 Komu je metodika určena? .....	4
1.2 Jak s metodikou pracovat?.....	5
2. Cíl metodiky.....	6
3. Popis metodiky.....	7
3.1 Podstata metodiky – obecný a specifický rámec metodiky.....	7
3.1.1 Ilustrativní příklad řešení modelových situací .....	9
3.2 Specifický rámec metodiky – podstata a název modelových situací.....	12
3.2.1 Spolehlivost (škály) ověřovacího testu.....	12
3.2.2 Zvládnutí testové položky žáky a stanovení obtížnosti testové položky.....	12
3.2.3 Schopnost testové položky rozlišit mezi žáky podle jejich výsledků v ověřovacím testu, identifikace matoucí správné odpovědi a potenciálně chybný klíč hodnocení.....	13
3.2.4 Kvalita nabízených nesprávných odpovědí testové položky.....	13
3.2.5 Změna spolehlivosti ověřovacího testu při vynechání testové položky.....	14
3.2.6 (Ne)spravedlnost testových položek k různým skupinám žáků .....	14
3.2.7 Kvalita testových položek a identifikace nekvalitních testových položek.....	15
3.2.8 Neobvyklý vzor odpovědí žáků na testové položky.....	15
3.2.9 Volba a využití škály pro stanovení výsledků žáků v ověřovacím testu.....	16
3.2.10 Propojení výsledků žáků ve dvou ověřovacích testech se společnými (kotvícími) testovými položkami na stejnou škálu .....	16
3.2.11 Pokrok ve vzdělávání s využitím ověřovacích testů .....	17
3.2.12 Unidimenzionalita ověřovacího testu a počet konstruktů v něm obsažených....	17
3.2.13 Lokální nezávislost testových položek.....	18
3.2.14 Dílčí konstrukty ověřovacího testu .....	18

3.2.15	Volba vhodného modelu vycházejícího z IRT pro vyhodnocení ověřovacího testu .....	19
3.2.16	Základní vyhodnocení a reporting výsledků ověřovacího testu.....	19
3.2.17	Rozdíly ve výsledcích žáků dané rozdíly uvnitř škol a rozdíly mezi školami ...	20
3.2.18	Faktory vztahující se k výsledkům žáků v ověřovacím testu.....	20
3.3	Specifický rámec metodiky – řešení modelových situací .....	21
3.3.1	Spolehlivost (škály) ověřovacího testu.....	21
3.3.2	Zvládnutí testové položky žáky a stanovení obtížnosti testové položky.....	23
3.3.3	Schopnost testové položky rozlišit mezi žáky podle jejich výsledků v ověřovacím testu, identifikace matoucí správné odpovědi a potenciálně chybný klíč hodnocení.....	24
3.3.4	Kvalita nabízených nesprávných odpovědí testové položky.....	26
3.3.5	Změna spolehlivosti ověřovacího testu při vynechání testové položky .....	28
3.3.6	(Ne)spravedlnost testových položek k různým skupinám žáků .....	29
3.3.7	Kvalita testových položek a identifikace nekvalitních testových položek.....	32
3.3.8	Neobvyklý vzor odpovědí žáků na testové položky.....	35
3.3.9	Volba a využití škály pro stanovení výsledků žáků v ověřovacím testu.....	38
3.3.10	Propojení výsledků žáků ve dvou ověřovacích testech se společnými (kotvícími) testovými položkami na stejnou škálu .....	43
3.3.11	Pokrok ve vzdělávání s využitím ověřovacích testů .....	47
3.3.12	Unidimenzionalita ověřovacího testu a počet konstruktů v něm obsažených....	48
3.3.13	Lokální nezávislost testových položek.....	52
3.3.14	Dílčí konstrukty ověřovacího testu .....	54
3.3.15	Volba vhodného modelu vycházejícího z IRT pro vyhodnocení ověřovacího testu .....	55
3.3.16	Základní vyhodnocení a reporting výsledků ověřovacího testu.....	57
3.3.17	Rozdíly ve výsledcích žáků dané rozdíly uvnitř škol a rozdíly mezi školami ...	58
3.3.18	Faktory vztahující se k výsledkům žáků v ověřovacím testu.....	59

4.	Inovativnost metodiky.....	61
5.	Popis uplatnění metodiky.....	62
6.	Seznam odkazované literatury.....	63

## 1. Úvod

Hodnocení a zpětná vazba jsou významnými faktory, které ovlivňují kvalitu počátečního vzdělávání žáků. Z řady existujících forem a metod hodnocení výsledků ve vzdělávání patří k nejvíce rozšířeným ověřovací testy využívané v řadě aplikačních situací, jako je například hodnocení výsledků na úrovni vzdělávacího systému, realizace přijímacích zkoušek, rozhodnutí o udělení profesního certifikátu, nastavení podoby kurikula či zdůvodnění akontability veřejných financí alokovaných pro oblast vzdělávání (např. Ryan a Brockmann, 2009; Thompson, 2015; Cook a Eignor, 1991; Lambert et al., 2018; Betebenner a Linn, 2010; Udofia a Uko, 2016; Betebenner, 2009). Ne náhodou klade na oblast hodnocení vysoký důraz také hlavní koncepční dokument České republiky pro oblast vzdělávání *Strategie vzdělávací politiky České republiky do roku 2030+* (Fryč et al., 2020).

Pro využití potenciálu ověřovacích testů ve vzdělávání byla navržena široká, ne vždy však plně využívaná, nabídka teoreticko-metodických přístupů k jejich vyhodnocení (Thompson, 2016). Taková praxe má ovšem nežádoucí vliv na uchopení příležitostí spojených s vyhodnocením ověřovacích testů a je hlavním motivem pro vytvoření *Metodiky vyhodnocení výsledků ověřovacího testování v počátečním vzdělávání* (dále jen „metodika“). Metodika se v tomto ohledu zaměřuje na ověřovací testy, které jsou utvářeny testovými položkami, na něž žák odpovídá buď správně, nebo nesprávně (dichotomické testové položky). Podstata vyhodnocení ověřovacích testů obsahujících testové položky, které jsou hodnoceny na škále s více než dvěma hodnotami (polytomické testové položky), je analogická, nicméně konkrétní postupy je nezbytné přizpůsobit specifikům těchto testových položek.

### 1.1 Komu je metodika určena?

Metodika je určena všem aktérům, kteří využívají ověřovací testování v počátečním vzdělávání při své práci. Mezi tyto aktéry patří především organizace skutečně či potenciálně zajišťující ověřovací testování na úrovni vzdělávacího systému (např. Centrum pro zjišťování výsledků vzdělávání, Česká školní inspekce, místní akční skupiny), a dále pak školy či jiné organizace působící v počátečním vzdělávání, které ověřovací testy žáků připravují či vyhodnocují. Záměrem metodiky je poskytnout osobě, která je odpovědná za realizaci a vyhodnocení ověřovacího testu v počátečním vzdělávání, podpůrný nástroj v její práci.



## 1.2 Jak s metodikou pracovat?

Metodika je ze své podstaty metodikou podpůrnou, neboť neposkytuje svým uživatelům jediný *one-size-fits-all* přístup, který by měl být sledován za každé situace. Metodika naopak poskytuje návodné postupy, které její uživatel volí a sleduje s ohledem na své vlastní záměry a předpoklady hodnocení. Rozhodnutí o podobě využití metodiky tak přísluší uživateli.

Koncepčně je metodika založena na typových situacích, které jsou označeny jako situace modelové. Pro všechny modelové situace je formulován společný, tzv. obecný rámec, který je v navazujícím rámci specifickém rozveden pro každou modelovou situaci zvlášť. Součástí obecného rámce modelových situací je podstata modelové situace, jejíž rozvedení ve specifickém rámci slouží uživateli metodiky k výběru modelových situací, které jsou vhodné pro naplnění jeho záměru hodnocení. Specifický rámec rovněž uvádí návodné postupy pro řešení modelových situací.

Navrhované metodické postupy pro řešení modelových situací vyžadují využití softwarové podpory. V tomto ohledu je metodika postavena na tzv. balíčcích (knihovnách, *packages*) programovacího jazyka R (dále jen „R balíček“), přičemž tato volba je motivována:

- výbornými vlastnostmi jazyka R pro provádění pokročilých statistických analýz dat;
- snadnou dostupností R balíčků v rámci tzv. svobodné licence.

Uvedené motivy umožňují prakticky každému zájemci využití metodiky bez ohledu na jeho finanční či jiná omezení.

## 2. Cíl metodiky

**Cílem metodiky je poskytnout uživatelům komplexní podpůrný nástroj s návodnými metodickými postupy vyhodnocení ověřovacích testů v počátečním vzdělávání.**

Cíl metodiky je zasazen do následujících souvislostí:

- Metodika se zaměřuje na ověřovací testy, které jsou utvářeny pouze testovými položkami, na něž žák odpovídá buď správně, nebo nesprávně (dichotomické testové položky).
- Metodika představuje podpůrný nástroj, tj. nástroj na podporu aplikace v něm uvedených postupů při vyhodnocování ověřovacích testů v počátečním vzdělávání. Takto metodika nestanovuje jediný závazný postup vyhodnocování, naopak o konečné aktivaci celé metodiky či jen části metodiky rozhoduje uživatel.
- Metodika neposkytuje uzavřený a vyčerpávající přehled všech postupů při vyhodnocování ověřovacích testů v počátečním vzdělávání. Metodika naopak ponechává otevřený prostor jednak pro aktualizaci existujících postupů, jednak pro doplňování dalších v metodice neuvedených postupů.
- Metodika pracuje s pojmem konstrukt, který je chápán ve smyslu latentní proměnné měřené ověřovacím testem (např. úroveň gramotností žáků). V souladu s terminologií metodiky se ověřovací test skládá z dílčích testových položek.

### 3. Popis metodiky

Tato kapitola je popisem metodiky, kdy první podkapitola charakterizuje podstatu metodiky a druhá podkapitola pak rozvádí její specifický rámec pro jednotlivé modelové situace.

#### 3.1 Podstata metodiky – obecný a specifický rámec metodiky

Podstata metodiky je založena na teoreticko-metodických východiscích pro vyhodnocení ověřovacího testování v počátečním vzdělávání, přičemž tato východiska metodika převádí do aplikační roviny v podobě návodných postupů pro řešení tzv. **modelových situací**. Pro modelové situace metodika vymezuje dva dílčí rámce: (a) obecný rámec metodiky společný pro všechny modelové situace; a (b) specifický rámec metodiky, který je pro každou modelovou situaci jedinečný.

**Obecný rámec metodiky** určuje společný obsah pro rozvedení (specifikaci) všech modelových situací, přičemž obsahuje následující dílčí elementy:

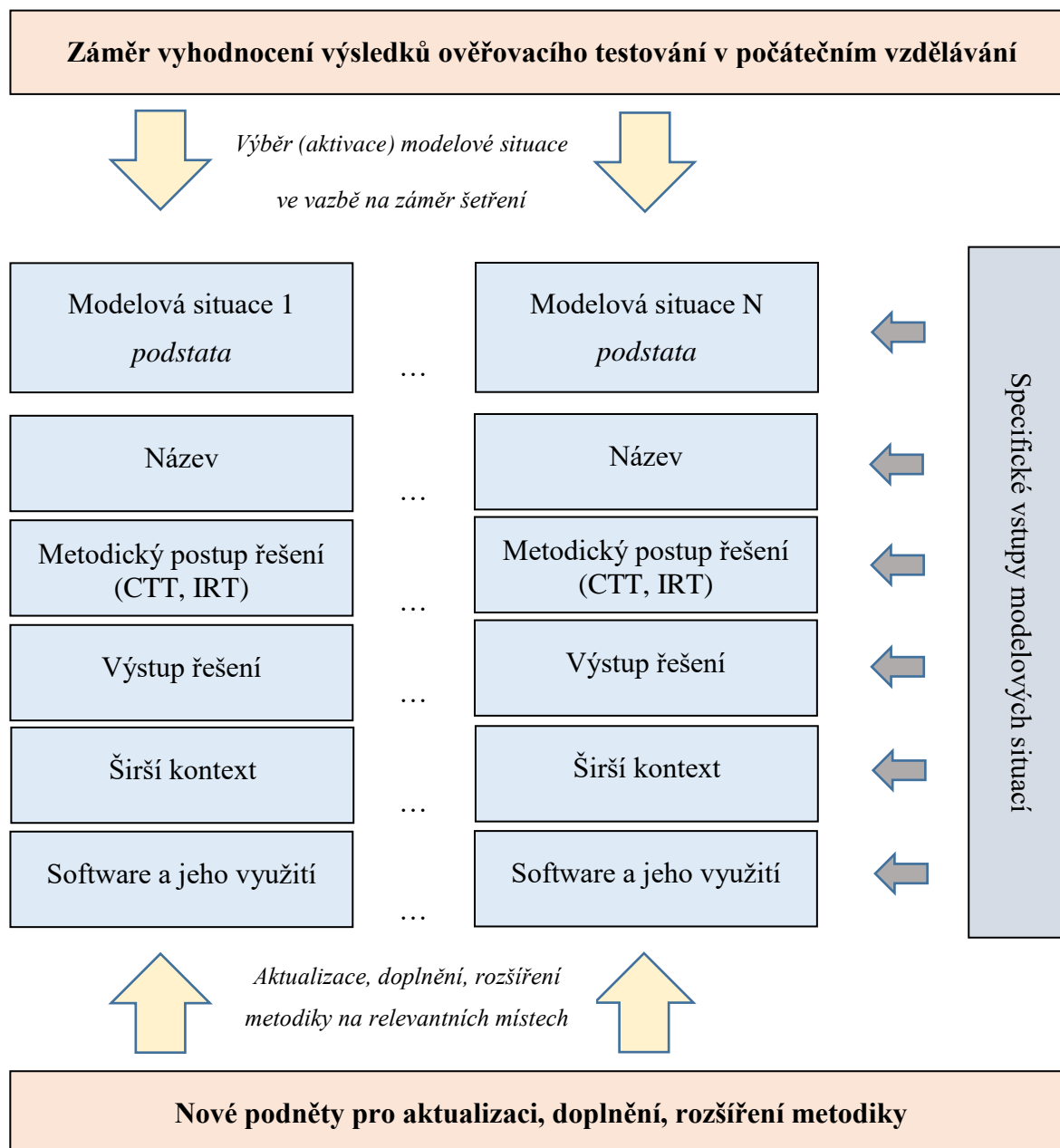
- název modelové situace;
- vysvětlení podstaty modelové situace;
- charakteristiku metodického postupu řešení modelové situace, který vychází z klasické teorie testů (dále jen „CTT“) nebo z teorie odpovědi na položku (dále jen „IRT“);
- představení výstupu řešení modelové situace;
- zasazení řešení modelové situace do širšího situačního kontextu (implikace a vazby k dalším modelovým situacím);
- uvedení relevantního software a ilustrace jeho využití při řešení modelové situace.

**Specifický rámec metodiky** respektuje jedinečnost každé modelové situace a potřebu rozvést (specifikovat) dílčí elementy obecného rámce pro každou modelovou situaci zvlášť. Takto specifický rámec vymezuje tu část metodiky, která je pro každou modelovou situaci jedinečná.

Podstata metodiky dále sleduje dva horizontální principy: (a) **princip výběru modelové situace ve vazbě na záměr vyhodnocení**; a (b) **princip otevřenosti metodiky** vůči novým podnětům. První z těchto principů je spojený s aktivací řešení modelových situací podle záměru uživatele metodiky. V souladu s tímto principem je uživatelem metodiky aktivován odpovídající počet modelových situací, de facto od jedné až po všechny z nich. Při výběru modelových situací k aktivaci posuzuje uživatel metodiky především soulad svého záměru

s podstatou modelových situací. Princip otevřenosti metodiky pak respektuje potřebu její flexibility ve vazbě jak na vývoj vědeckého poznání, tak na utváření nových potřeb souvisejících s ověřovacím testováním v počátečním vzdělávání. V souladu s principem otevřenosti tak je možné metodiku aktualizovat a doplňovat v reakci na nové podněty. Schematicky je podstata metodiky zachycena na obrázku č. 1.

**Obrázek č. 1: Schéma podstaty metodiky**



Podstata metodiky uznává význam předpokladů pro správnou aplikaci metod vycházejících z CTT a především z IRT, přičemž ověřování naplnění těchto předpokladů je součástí řešení souvisejících modelových situací. V obecné rovině pak metodika především zdůrazňuje potřebu respektovat požadavky kladené na velikost výběrového souboru žáků řešících ověřovací test pro aktivaci metodických postupů modelových situací, které jsou založeny na odhadech modelů vycházejících z IRT. Návodná doporučení v tomto ohledu jsou: (a) minimální počet 100 až 200 žáků pro odhady 1PL modelu (např. DeMars, 2010; Thorpe a Favia, 2012; Edelen a Reeve, 2007); (b) minimální počet 500 žáků pro odhady 2PL modelu (např. De Champlain, 2010; DeMars, 2010; Edelen a Reeve, 2007); a (c) minimální počet až 1000 žáků pro odhady 3PL modely (např. DeMars, 2010).

### 3.1.1 Ilustrativní příklad řešení modelových situací

Dílčím elementem obecného rámce metodiky je uvedení relevantního software a ilustrace jeho využití při řešení modelových situací. Naplnění tohoto záměru je ve specifickém rámci metodiky založeno na funkcích uložených v R balíčcích, které je potřeba aktivovat v grafickém uživatelském rozhraní, v němž uživatel metodiky s programovacím jazykem R pracuje (blíže viz Venables et al., 2020). Pro ilustraci řešení modelových situací jsou dále využity následující datové soubory:

- Datový soubor s označením *TACR\_DATA.csv* je základní test, který obsahuje odpovědi tři tisíc žáků na 38 dichotomických testových položek (*item1*, *item2*, ..., *item38*) nabízejících výběr ze čtyř odpovědí. Správné odpovědi žáků na testové položky jsou kódovány hodnotou „1“, nesprávné odpovědi žáků pak hodnotou „0“.
- Datový soubor s označením *TACR\_DATA\_DISTRAKT.csv* obsahuje všechny testové položky základního testu a navíc šest proměnných (*item27A*, *item27B*, *item27C*, *item32A*, *item32C* a *item32D*). Proměnná *item27A* nabývá hodnotu „1“, pokud žák vybral nesprávnou odpověď A jako své řešení testové položky 27, v ostatních případech pak hodnotu „0“. Interpretace zbývajících pěti dodatečných proměnných je analogická.
- Datový soubor s označením *TACR\_DATA\_DIF.csv* obsahuje všechny testové položky základního testu a dále proměnnou *divka*, která nabývá hodnotu „1“, pokud na danou testovou položku odpovídala dívka, v opačném případě pak hodnotu „0“.
- Datový soubor s označením *TACR\_DATA\_ALTER.csv* je test alternativní k základnímu testu. Stejně jako základní test obsahuje i alternativní test odpovědi tři tisíc žáků

na 38 dichotomických testových položek nabízejících výběr ze čtyř odpovědí, výběrové soubory žáků řešící jednotlivé testy jsou však odlišné. Základní a alternativní test mají 20 testových položek společných (*item1*, *item2*, ..., *item20*), dalších 18 testových položek (*item39*, *item40*, ..., *item56*) je jedinečných pro druhý test. Opětovně platí, že hodnota „1“ vyjadřuje správnou odpověď žáka, hodnota „0“ pak nesprávnou odpověď žáka.

- Datový soubor s označením *TACR\_DATA\_EQUATE1.csv* obsahuje dvě proměnné pro každého žáka. Proměnná *total1* vyjadřuje počet žákem správně zodpovězených testových položek pro základní test (testové položky *item1*, ..., *item38*). Proměnná *anchor1* vyjadřuje počet žákem správně zodpovězených testových položek, které mají základní a alternativní test společné (*item1*, *item2*, ..., *item20*).
- Datový soubor s označením *TACR\_DATA\_EQUATE2.csv* obsahuje dvě proměnné pro každého žáka. Proměnná *total2* vyjadřuje počet žákem správně zodpovězených testových položek pro alternativní test (testové položky *item1*, ..., *item20*, *item39*, ..., *item56*). Proměnná *anchor2* vyjadřuje počet žákem správně zodpovězených testových položek, které mají základní a alternativní test společné (*item1*, *item2*, ..., *item20*).
- Datový soubor s označením *TACR\_DATA\_HIEARCH.csv* obsahuje šest proměnných, a to:  
(a) *IDskoly* s číselným označením od 1 do 120; (b) *uspesnost* odpovídající podílu žákem správně zodpovězených testových položek základního testu; (c) *body* odpovídající výsledku žáka v základním testu na bodové škále se středem 500 bodů a směrodatnou odchylkou 100 bodů odvozené z odhadu 3PL modelu; (d) *skore* odpovídající počtu žákem správně zodpovězených testových položek základního testu; (e) *divka* nabývající hodnotu „1“, pokud na danou testovou položku odpovídala dívka, v opačném případě pak hodnotu „0“; a (f) *vyssi\_soceko\_status* nabývající hodnotu „1“, pokud žák pochází z rodinného zázemí s vyšším socioekonomickým statusem, v opačném případě pak hodnotu „0“.

Všechny výše uvedené datové soubory, které jsou využity pro ilustraci řešení modelových situací, jsou na vyžádání k dispozici u hlavního řešitele projektu.

Pro možnost pracovat s uvedenými datovými soubory v programovacím jazyce R je potřeba jejich načtení v grafickém uživatelském rozhraní. Tohoto záměru je možné dosáhnout různými způsoby, dále uvedený postup je jedním z nich a je možné jej využít pro řešení všech modelových situací.

**Načtení datových souborů pro práci v programovacím jazyce R**

```
setwd("C:/TACR/Metodika")
```

*# Funkce setwd určuje pracovní adresář, odkud jsou nahrávány a kam jsou ukládány pracovní soubory.*

```
TACR_DATA <- read.csv("TACR_DATA.csv", header = TRUE, sep = ";", dec = ",")
```

*# Funkce read.csv umožňuje načíst csv datový soubor základního testu (TACR\_DATA.csv) jako datový rámec TACR\_DATA. Pro definici objektů je v programovacím jazyce R užíván operátor <-.*

*# Atribut header vyjadřuje, že datový soubor obsahuje označení proměnných v záhlaví.*

*# Atribut sep vyjadřuje, že data jsou v csv datovém souboru oddělena středníkem.*

*# Atribut dec vyjadřuje podobu desetinné čárky.*

```
attach(TACR_DATA)
```

*# Funkce attach umožňuje pracovat s jednotlivými proměnnými datového rámce TACR\_DATA.*

```
str(TACR_DATA)
```

*# Funkce str umožňuje zobrazit proměnné datového rámce TACR\_DATA a jejich základní charakteristiky.*

### 3.2 Specifický rámec metodiky – podstata a název modelových situací

Ve specifickém rámci jsou pro jednotlivé modelové situace rozvedeny (specifikovány) dílčí elementy obecného rámce metodiky. Primárně je v této podkapitole vysvětlena podstata modelových situací tak, aby uživatel metodiky mohl snadněji vybrat vhodnou modelovou situaci pro svůj záměr vyhodnocení ověřovacího testu. Po výběru modelové situace přechází uživatel metodiky k návodnému postupu jejího řešení v podkapitole 3.3.

#### 3.2.1 Spolehlivost (škály) ověřovacího testu

Název modelové situace
Spolehlivost (škály) ověřovacího testu
Podstata modelové situace
Podstata modelové situace reaguje na záměr uživatele metodiky zjistit, jaká je spolehlivost (škály) ověřovacího testu. Spolehlivost (škály) ověřovacího testu může být negativně ovlivněna řadou vlivů, jako jsou: (a) nekontrolované podmínky testování (např. prostorové a časové podmínky testování); (b) náhodná fluktuace výsledků žáka spojená s jeho osobními náladami a charakteristikami; a (c) chyba spojená s kvalitou slovního vyjádření testových položek. Zájmem uživatele metodiky je dosáhnout vysoké spolehlivosti ověřovacího testu, přičemž tento zájem se zvyšuje v závislosti na důležitosti, tj. praktických dopadech, ověřovacího testu.

#### 3.2.2 Zvládnutí testové položky žáky a stanovení obtížnosti testové položky

Název modelové situace
Zvládnutí testové položky žáky a stanovení obtížnosti testové položky
Podstata modelové situace
Podstata modelové situace reaguje na záměr uživatele metodiky poznat ty testové položky, které žáci zvládají dobře a ty testové položky, které žáci zvládají hůře. V souladu s tímto záměrem je zájem uživatele metodiky posoudit obtížnost jednotlivých testových položek ověřovacího testu, což je také jeden z důležitých ukazatelů kvality testových položek.



### 3.2.3 Schopnost testové položky rozlišit mezi žáky podle jejich výsledků v ověřovacím testu, identifikace matoucí správné odpovědi a potenciálně chybný klíč hodnocení

Název modelové situace
Schopnost testové položky rozlišit mezi žáky podle jejich výsledků v ověřovacím testu, identifikace matoucí správné odpovědi a potenciálně chybný klíč hodnocení
Podstata modelové situace
Základní podstata modelové situace reaguje na záměr uživatele metodiky zjistit, jak dobře testová položka rozlišuje mezi žáky, kteří dosáhli dobrého výsledku v ověřovacím testu a žáky, kteří dosáhli slabého výsledku v ověřovacím testu. Tento záměr je motivován zájmem uživatele metodiky o poznání další indicie kvality testové položky, tj. její schopnosti rozlišit mezi žáky s dobrými a slabými výsledky v ověřovacím testu.  Doplňující podstata modelové situace je motivována záměrem ověřit, zda: (a) není správná odpověď testové položky formulována matoucím, či pochybným způsobem; a (b) je klíč k řešení testových položek ověřovacího testu správný. Přirozeným je v tomto ohledu předpoklad, že žáci s dobrými výsledky zodpovídají testovou položku správně s vyšší pravděpodobností než žáci se slabými výsledky v ověřovacím testu. Opačné zjištění může naznačovat jeden z uvedených problémů.

### 3.2.4 Kvalita nabízených nesprávných odpovědí testové položky

Název modelové situace
Kvalita nabízených nesprávných odpovědí testové položky
Podstata modelové situace
Podstata modelové situace je relevantní pro testové položky, kdy žák vybírá odpověď z konečné nabídky možností. Záměrem uživatele metodiky je posoudit kvalitu nabízených nesprávných odpovědí (distraktorů), především zda: (a) některá z nesprávných odpovědí neláká pozornost především dobrých žáků; a (b) některá z nesprávných odpovědí není zbytečná, protože ji nevolí prakticky žádný žák. Obecně pak modelová situace poskytuje další informace o kvalitě testové položky, a to ve vazbě na formulaci nabídky nesprávných odpovědí (distraktorů).

### 3.2.5 Změna spolehlivosti ověřovacího testu při vynechání testové položky

<b>Název modelové situace</b>
Změna spolehlivosti ověřovacího testu při vynechání testové položky
<b>Podstata modelové situace</b>
Podstata modelové situace reaguje na záměr uživatele metodiky posoudit, zda se spolehlivost ověřovacího testu změní při vynechání testové položky. Tento záměr je motivován zájmem uživatele metodiky posoudit kvalitu testové položky vzhledem k hodnocenému konstrukt. Pokud testová položka snižuje spolehlivost (škály) ověřovacího testu, může uživatel metodiky rozhodnout o případném vynechání takové testové položky z vyhodnocení.

### 3.2.6 (Ne)spravedlnost testových položek k různým skupinám žáků

<b>Název modelové situace</b>
(Ne)spravedlnost testových položek k různým skupinám žáků
<b>Podstata modelové situace</b>
Podstata modelové situace reaguje na záměr uživatele metodiky posoudit, zda se chování (charakteristiky, parametry) testových položek neliší vzhledem k různým skupinám žáků (např. chlapci a dívky). Taková situace vzniká, když dvě skupiny žáků se stejnou úrovní zvládnutí hodnoceného konstrukt zodpovídají testovou položku odlišně, a testová položka se pak nechová spravedlivě. Rozlišit přitom lze:  (a) uniformní rozdíly mezi skupinami žáků, které jsou zaznamenány pro všechny úrovně zvládnutí hodnoceného konstrukt;  (b) neuniformní rozdíly mezi skupinami žáků, které jsou zaznamenány jen pro žáky na určité úrovni zvládnutí hodnoceného konstrukt.  Rozdíly mezi skupinami žáků, tj. nespravedlnost testové položky, mohou být způsobeny například formulací jejího zadání, které může být bližší a tím i lépe srozumitelné jedné skupině žáků a vzdálenější a tím i hůře srozumitelné druhé skupině žáků. Zároveň se přitom vyžaduje, aby se téma zadání testové položky přímo nevztahovalo k testem hodnocenému konstrukt.

### 3.2.7 Kvalita testových položek a identifikace nekvalitních testových položek

Název modelové situace
Kvalita testových položek a identifikace nekvalitních testových položek
Podstata modelové situace
Podstata modelové situace reaguje na záměr uživatele metodiky komplexně posoudit kvalitu testových položek ověřovacího testu. Tento záměr je především motivován zájmem uživatele metodiky nalézt nekvalitní testové položky, tj. testové položky s nežádoucími charakteristikami, které se především týkají: (a) jejich obtížnosti; (b) jejich schopnosti rozlišit mezi žáky s dobrými a slabými výsledky v ověřovacím testu; (c) kvality nabízených nesprávných odpovědí (distraktorů); (d) změny spolehlivosti ověřovacího testu při jejich vynechání; a (e) jejich (ne)spravedlnosti k různým skupinám žáků. V případě modelů vycházejících z IRT je dále posuzována dobrá shoda skutečných a modelem predikovaných odpovědí žáků na testovou položku. V případě identifikovaných nekvalitních testových položek uživatel metodiky rozhoduje o způsobu jejich vyhodnocení. Modelová situace pak také poskytuje vstupní informace o testových položkách při rozhodování o jejich využití při tvorbě nového ověřovacího testu.

### 3.2.8 Neobvyklý vzor odpovědí žáků na testové položky

Název modelové situace
Neobvyklý vzor odpovědí žáků na testové položky
Podstata modelové situace
Podstata modelové situace reaguje na záměr uživatele metodiky identifikovat ty žáky, jejichž odpovědi na testové položky ověřovacího testu lze považovat za neobvyklé až podivné, neboť žáci odpovídají správně na obtížné testové položky a naopak nesprávně na jednoduché testové položky. Podivný vzor odpovědí žáků na testové položky ověřovacího testu může vznikat například v důsledku: (a) neetického chování žáka; (b) nízké motivace žáka vedoucí k hádání odpovědí na testové položky; a (c) nedostatečných znalostí a dovedností žáka v některé z testovaných oblastí. V případě identifikace neobvyklých odpovědí žáků na testové položky rozhoduje uživatel metodiky o tom, zda budou odpovědi těchto žáků zahrnuty do vyhodnocení ověřovacího testu či nikoliv.

### 3.2.9 Volba a využití škály pro stanovení výsledků žáků v ověřovacím testu

<b>Název modelové situace</b>
Volba a využití škály pro stanovení výsledků žáků v ověřovacím testu
<b>Podstata modelové situace</b>
Podstata modelové situace reaguje na základní záměr ověřovacího testování – získat informaci o úrovni zvládnutí hodnoceného konstruktů žáky, kteří se ověřovacího testování účastní. Záměrem uživatele metodiky v modelové situaci proto je: (a) zvolit škálu, na které bude úroveň zvládnutí hodnoceného konstruktů žáky hodnocena; a (b) vyhodnotit úroveň zvládnutí hodnoceného konstruktů žáky na zvolené škále (využití škály). Uživateli metodiky se při volbě škály nabízí řada možností, které mají své specifické předpoklady využití.

### 3.2.10 Propojení výsledků žáků ve dvou ověřovacích testech se společnými (kotvícími) testovými položkami na stejnou škálu

<b>Název modelové situace</b>
Propojení výsledků žáků ve dvou ověřovacích testech se společnými (kotvícími) testovými položkami na stejnou škálu
<b>Podstata modelové situace</b>
Podstata modelové situace reaguje na situaci, kdy dva neekvivalentní výběrové soubory žáků (vzhledem k populaci i úrovni zvládnutí hodnoceného konstruktů) řeší různé verze ověřovacího testu, přičemž však existuje množina testových položek, která je pro ně společná (kotvící testové položky). Taková modelová situace může například nastat v případě záměru uživatele metodiky:  (a) propojit různé verze ověřovacího testu v daném roce a ročníku;  (b) propojit starší a novější verze ověřovacího testu v různých letech testování;  (c) utvořit škálu měřící úroveň zvládnutí hodnoceného konstruktů žáky různých ročníků;  (d) propojit verze ověřovacího testu národních a mezinárodních šetření.  Zájmem uživatele metodiky je propojit výsledky žáků řešících odlišné ověřovací testy se společnými (kotvícími) testovými položkami na stejnou škálu tak, aby žáci řešící obtížnější verzi testu nebyli znevýhodněni.

### 3.2.11 Pokrok ve vzdělávání s využitím ověřovacích testů

Název modelové situace
Pokrok ve vzdělávání s využitím ověřovacích testů
Podstata modelové situace
Podstata modelové situace reaguje na záměr uživatele metodiky posoudit pokrok ve vzdělávání, tj. změnu vzdělávacích výsledků v čase, prostřednictvím ověřovacích testů. Hodnocení se může týkat různých úrovní: (a) žák; (b) třída (učitel); a (c) škola. Základním předpokladem modelové situace je dostupnost dat z ověřovacích testů pro stejné žáky a za více období tak, aby bylo možné změnu ve vzdělávacích výsledcích žáků vyhodnotit.

### 3.2.12 Unidimenzionalita ověřovacího testu a počet konstruktů v něm obsažených

Název modelové situace
Unidimenzionalita ověřovacího testu a počet konstruktů v něm obsažených
Podstata modelové situace
Podstata modelové situace vychází ze skutečnosti, že unidimenzionalita je jedním z hlavních předpokladů odhadů řady statistik a indexů spojených s vyhodnocením ověřovacích testů, především pak tradičních modelů vycházejících z IRT. Další konstrukty (dimenze, faktory) se mohou v ověřovacím testu vyskytovat například v situacích, kdy:  (a) test není tvořen jedním hlavním konstruktem, ale dvěma odlišnými konstrukty, u nichž teorie předpokládala spojení v podobě hlavního konstrukt;  (b) testové položky uchopují doplňující konstrukty související s konstruktem hlavním (např. čtenářská gramotnost žáků daná delším uvozujícím textem k matematickým úlohám);  (c) žáci odpovídají na daný konstrukt odlišně vlivem svých psychologických procesů, jako je například motivace;  (d) je v datech přítomen faktor příslušnosti ke skupině žáků, tj. testové položky nejsou spravedlivé vůči různým skupinám žáků.  Záměrem uživatele metodiky je proto posoudit, zda ověřovací test měří jen jeden hlavní konstrukt (dimenzi, faktor) a pokud nikoliv, stanovit počet konstruktů v testu obsažených a současně sílu konstruktů hlavního.

### 3.2.13 Lokální nezávislost testových položek

Název modelové situace
Lokální nezávislost testových položek
Podstata modelové situace
<p>Podstata modelové situace vychází ze skutečnosti, že lokální nezávislost testových položek je jedním z hlavních předpokladů odhadů řady statistik a indexů spojených s vyhodnocením ověřovacích testů, především pak tradičních modelů vycházejících z IRT. Lokální nezávislost testových položek může být narušena například z důvodů:</p> <ul style="list-style-type: none"><li>(a) jejich podobného obsahu;</li><li>(b) jejich odvození ze stejného uvozujícího textu;</li><li>(c) jejich vzájemného propojení a podmíněnosti odpovědí.</li></ul> <p>Záměrem uživatele metodiky je proto posoudit, zda je lokální závislost testových položek v ověřovacím testu přítomna a jaké podoby nabývá.</p>

### 3.2.14 Dílčí konstrukty ověřovacího testu

Název modelové situace
Dílčí konstrukty ověřovacího testu
Podstata modelové situace
<p>Podstata modelové situace reaguje na záměr uživatele metodiky poznat vazby mezi testovými položkami, tj. vnitřní strukturu testových položek ověřovacího testu a na tomto základě zjistit, jaké dílčí konstrukty ověřovací test obsahuje. Podstatou modelové situace tedy je nalézt vzájemně související testové položky a pojmenovat dílčí konstrukty, které související testové položky utvářejí. Pokud byly dílčí konstrukty ověřovacího testu expertně definovány před testováním (např. algebraické dovednosti a geometrické dovednosti jako dílčí dimenze matematické gramotnosti), umožňuje modelová situace navíc ověřit kvalitu expertně definovaných témat ověřovacího testu, kdy otázkou je, zda zvolené testové položky skutečně měří definovaný konstrukt.</p>

### 3.2.15 Volba vhodného modelu vycházejícího z IRT pro vyhodnocení ověřovacího testu

<b>Název modelové situace</b>
Volba vhodného modelu vycházejícího z IRT pro vyhodnocení ověřovacího testu
<b>Podstata modelové situace</b>
<p>Podstata modelové situace vychází ze záměru uživatele metodiky vybrat vhodný model, který vychází z IRT, pro vyhodnocení ověřovacího testu. Postup spojený s výběrem takového modelu posuzuje, jak dobře model charakterizuje empirická data ověřovacího testu, respektive jak dobře mohly být odpovědi žáků na testové položky ověřovacího testu vygenerovány zvažovaným modelem. Žádoucí je v tomto ohledu vysoká úroveň shody dat ověřovacího testu a modelem generovaných dat. Modelovou situaci lze rovněž rozšířit na výběr nejvhodnějšího modelu vycházejícího z IRT pro vyhodnocení ověřovacího testu, tj. modelu, který vykazuje nejvyšší úroveň shody empirických a modelem generovaných dat. Pokud mezi dvěma modely není v tomto ohledu identifikován významný rozdíl, je preferován jednodušší model.</p>

### 3.2.16 Základní vyhodnocení a reporting výsledků ověřovacího testu

<b>Název modelové situace</b>
Základní vyhodnocení a reporting výsledků ověřovacího testu
<b>Podstata modelové situace</b>
<p>Podstata modelové situace reaguje na záměr uživatele metodiky vyhodnotit a představit výsledky ověřovacího testu, a to pro některou z následujících situací: (a) vyhodnocení a reporting výsledků ověřovacího testu vůči dílčím tematickým částem ověřovacího testu; (b) vyhodnocení a reporting výsledků ověřovacího testu na úrovni žáka, třídy/učitele, školy, území a systému; (c) vyhodnocení a reporting výsledků ověřovacího testu vůči dílčím charakteristikám žáka, třídy/učitele, školy a území. Pro řešení modelové situace je úzce souvisejícím úkolem volba škály vyhodnocení a reportingu dosažené úrovně hodnoceného konstruktů ověřovacího testu. Důležitým předpokladem modelové situace pak je dostupnost informací o tematických částech ověřovacího testu, respektive o charakteristikách žáků, tříd/učitelů, škol a území.</p>

**3.2.17 Rozdíly ve výsledcích žáků dané rozdíly uvnitř škol a rozdíly mezi školami**

<b>Název modelové situace</b>
Rozdíly ve výsledcích žáků dané rozdíly uvnitř škol a rozdíly mezi školami
<b>Podstata modelové situace</b>
Podstata modelové situace reaguje na záměr uživatele metodiky posoudit, do jaké míry jsou rozdíly ve výsledcích žáků v ověřovacím testu utvářeny: (a) rozdíly ve výsledcích žáků uvnitř škol; a (b) rozdíly ve výsledcích žáků mezi školami. Vysoký vliv rozdílů ve výsledcích žáků mezi školami naznačuje více selektivní charakter škol, nízký vliv rozdílů ve výsledcích žáků mezi školami naznačuje vyšší význam rovného přístupu ke vzdělání na úrovni škol.

**3.2.18 Faktory vztahující se k výsledkům žáků v ověřovacím testu**

<b>Název modelové situace</b>
Faktory vztahující se k výsledkům žáků v ověřovacím testu
<b>Podstata modelové situace</b>
Podstata modelové situace reaguje na záměr uživatele metodiky posoudit, které faktory mají významný vztah k výsledkům žáků v ověřovacím testu. Množina takových faktorů je velmi široká a může se týkat různých úrovní: (a) žáka (např. charakteristiky a postoje žáků); (b) třídy/učitele (např. charakteristiky výuky); (c) školy (např. charakteristiky materiálních a finančních podmínek školy); a (d) území (např. charakteristiky obyvatelstva území). Při hodnocení vztahu faktorů k výsledkům žáků v ověřovacím testu je potřeba vzít vliv různých úrovní do úvahy.



### 3.3 Specifický rámec metodiky – řešení modelových situací

V této podkapitole je podstata modelových situací rozvedena o jejich řešení, kdy uživatel metodiky volí příslušnou kartu v návaznosti na název jim zvolené modelové situace.

#### 3.3.1 Spolehlivost (škály) ověřovacího testu

Název modelové situace	
Spolehlivost (škály) ověřovacího testu	
Metodický postup řešení modelové situace	
<p>Metodický postup řešení modelové situace, která hodnotí spolehlivost (škály) ověřovacího testu na základě odpovědí žáků na testové položky, se skládá z následujících kroků uživatele metodiky:</p> <p>Krok 1: Uživatel metodiky vybírá, zda bude spolehlivost ověřovacího testu hodnotit s využitím odhadu modelu vycházejícího z IRT.</p>	
<i>Ne (přístup CTT)</i>	<i>Ano (přístup IRT)</i>
<p>Krok 2: Uživatel metodiky počítá hodnoty šesti ukazatelů spolehlivosti testu (<math>\lambda_1</math> až <math>\lambda_6</math>), kdy <math>\lambda_3</math> je Cronbachovo alfa.</p> <p>Krok 3: Uživatel metodiky počítá dva ukazatele spolehlivosti testu <math>\omega_h</math> a <math>\omega_t</math>, kde ukazatel <math>\omega_h</math> se vztahuje jen k hlavnímu konstruktů a ukazatel <math>\omega_t</math> zohledňuje také vliv skupinových konstruktů dílčích testových položek.</p>	<p>Krok 2: Uživatel metodiky odhaduje zvolený model v souladu s metodickým postupem modelové situace „Volba a využití škály pro stanovení výsledků žáků v ověřovacím testu“.</p> <p>Krok 3: Uživatel metodiky odvozuje a vykresluje informační křivku testu.</p>
Výstup řešení modelové situace	
<p>Výstupem řešení modelové situace je osm hodnot ukazatelů spolehlivosti <math>\lambda_1</math> až <math>\lambda_6</math>, <math>\omega_h</math> a <math>\omega_t</math> v intervalu od 0 do 1. Vyšší hodnoty ukazatelů spolehlivosti <math>\lambda_1</math> až <math>\lambda_6</math> a <math>\omega_h</math> ukazují na vyšší spolehlivost testu. Nižší hodnoty ukazatele spolehlivosti <math>\omega_h</math> a vysoké hodnoty ukazatele <math>\omega_t</math> ukazují na slabší vliv hlavního konstruktů a silný vliv skupinových konstruktů dílčích testových položek (vedlejší konstrukty ověřovacího testu). Taková situace narušuje důvěryhodnost ukazatelů spolehlivosti <math>\lambda_1</math> až <math>\lambda_6</math>.</p>	<p>Výstupem řešení modelové situace je vykreslení informační křivky testu, která na ose x zachycuje úroveň zvládnutí hodnoceného konstruktů a na ose y korespondující množství poskytnuté informace (spolehlivost) testu. Vyšší množství poskytnuté informace znamená vyšší spolehlivost testu na dané úrovni zvládnutí hodnoceného konstruktů žáky.</p>

## Širší situační kontext

Spolehlivost (škály) ověřovacího testu je logicky vztažena ke kvalitě testových položek, kdy vynechání nekvalitních testových položek typicky vede ke zvýšení spolehlivosti (škály) ověřovacího testu.

Nejnižší uváděná minimální hodnota Cronbachova alfa pro akceptovatelnou spolehlivost ověřovacího testu je 0,7, v případě testů s významnými dopady je však typicky vyžadována spolehlivost vyšší.

Srovnání hodnot ukazatelů spolehlivosti  $\omega_h$  a  $\omega_t$  je vhodným metodickým postupem pro posouzení unidimenzionality testu, a to ve vazbě na sílu vlivu hlavního konstruktů. Následně je možné pro detekci narušení předpokladu unidimenzionality využít další metodické postupy.

Informační křivka testu je vhodným nástrojem pro posouzení vhodnosti ověřovacího testu k záměru testování, neboť ukazuje množství poskytnuté informace na jednotlivých úrovních zvládnutí hodnoceného konstruktů žáky. Tam, kde je množství poskytnuté informace nedostatečné, lze přidat testovou položku vhodných charakteristik (parametry obtížnosti a diskriminace).

## Software a ilustrace jeho využití pro řešení modelové situace

R balíček *psych* (viz Revelle, 2020)

R balíček *ltm* (viz Rizopoulos, 2018)

## Přístup CTT pro datový rámec TACR\_DATA

```
guttman(TACR_DATA)
```

*# Funkce guttman (R balíček psych) umožňuje výpočet hodnot šesti Guttmanových ukazatelů  $\lambda_1$  až  $\lambda_6$ .*

```
omega(TACR_DATA)
```

*# Funkce omega (R balíček psych) umožňuje výpočet hodnot  $\omega_h$  a  $\omega_t$ .*

## Přístup IRT pro datový rámec TACR\_DATA a odhad 2PL modelu

```
PL2model <- ltm(TACR_DATA ~ z1)
```

*# Funkce ltm (R balíček ltm) vede k odhadu parametrů 2PL modelu, které jsou uloženy v objektu PL2model.*

```
plot(PL2model, type = "IIC", items = 0, xlab = "úroveň statistické gramotnosti", ylab = "informace", main = "")
```

*# Funkce plot (R balíček ltm) vykresluje informační křivku testu pro odhadovaný 2PL model z předchozího kroku metodického postupu.*

*# Atribut type s hodnotou IIC sděluje, že má být vykreslena informační křivka.*

*# Atribut items s hodnotou 0 sděluje, že má být vykreslena informační křivka celého testu, nejen jednotlivých testových položek. Pro vykreslení vybraných testových položek se zadává číslo jejich pořadí (např. items = c(1, 3, 7, 8)).*

*# Atribut xlab uvádí označení osy x, v tomto případě úroveň statistické gramotnosti.*

*# Atribut ylab uvádí označení osy y, v tomto případě informace.*

*# Atribut main uvádí název celého grafu, v tomto případě nebude zobrazen žádný název.*

## 3.3.2 Zvládnutí testové položky žáky a stanovení obtížnosti testové položky

Název modelové situace	
Zvládnutí testové položky žáky a stanovení obtížnosti testové položky	
Metodický postup řešení modelové situace	
Metodický postup řešení modelové situace je založen na výpočtu obtížnosti testové položky a skládá se z následujících kroků uživatele metodiky:	
Krok 1: Uživatel metodiky vybírá, zda bude obtížnost testové položky hodnotit s využitím odhadu modelu vycházejícího z IRT.	
<i>Ne (přístup CTT)</i>	<i>Ano (přístup IRT)</i>
Krok 2a: Uživatel metodiky počítá obtížnost testové položky jako podíl žáků účastnících se testu, kteří zodpověděli danou testovou položku správně.	Krok 2: Uživatel metodiky odhaduje zvolený model v souladu s metodickým postupem modelové situace „Volba a využití škály pro stanovení výsledků žáků v ověřovacím testu“.  Krok 3: Z odhadu zvoleného modelu uživatel metodiky extrahuje parametr obtížnosti testové položky.
Výstup řešení modelové situace	
Výstupem řešení modelové situace je hodnota obtížnosti pro dichotomické testové položky vyjádřená jako podíl žáků účastnících se testu, kteří zodpověděli danou testovou položku správně. Vysoké hodnoty tohoto podílu jsou spojeny s jednoduchými testovými položkami, naopak nízké hodnoty jsou spojeny s obtížnými testovými položkami.	Výstupem řešení modelové situace je hodnota parametru obtížnosti testové položky, která odpovídá úrovni hodnoceného konstruktů, na které přibližně 50 % testovaných žáků odpovídá testovou položku správně. Vysoké hodnoty tohoto parametru jsou spojeny s obtížnými testovými položkami, naopak nízké hodnoty jsou spojeny s jednoduchými testovými položkami.
Širší situační kontext	
<p>Obtížnost je jedním z klíčových parametrů charakterizujících kvalitu testové položky. Za málo přínosné jsou označovány testové položky s obtížností vyšší než 0,90, respektive nižší než 0,10 (0,20) při hodnocení podílu správně odpovídajících žáků (přístup CTT). V případě přístupů založených na IRT jsou jako podezřelé uváděny především testové položky s obtížností mimo interval -3 až +3.</p> <p>Obtížnost testové položky ovlivňuje spolehlivost testu na různých úrovních zvládnutí hodnoceného konstruktů, a to ve vazbě na množství poskytnuté informace, které je nejvyšší právě na úrovni obtížnosti testové položky.</p>	
Software a ilustrace jeho využití pro řešení modelové situace	
R balíček <i>CTT</i> (viz Willse, 2018)	R balíček <i>ltm</i> (viz Rizopoulos, 2018)

**Přístup CTT pro datový rámec TACR\_DATA**

```
Obtiznost_CTT <- itemAnalysis(TACR_DATA, itemReport=TRUE)
```

*# Funkce itemAnalysis (R balíček CTT) vede k výpočtu hodnot charakteristik kvality testových položek (včetně obtížnosti), které jsou uloženy v objektu Obtiznost\_CTT.*

*# Atribut itemReport uvádí, že v objektu Obtiznost\_CTT má být uložen také report obsahující hodnoty charakteristik kvality testových položek.*

```
Obtiznost_CTT$itemReport$itemMean
```

*# Příkaz vede k „vytažení“ reportu obsahujícího hodnoty charakteristik kvality testových položek (itemReport) a z nich pak „vytažení“ hodnoty obtížnosti testových položek (itemMean).*

**Přístup IRT pro datový rámec TACR\_DATA a odhad 3PL modelu**

```
PL3model <- tpm(TACR_DATA)
```

*# Funkce tpm (R balíček ltm) vede k odhadu parametrů 3PL modelu, které jsou uloženy v objektu PL3model.*

```
coef(PL3model)
```

*# Funkce coef (R balíček ltm) zobrazí hodnoty parametrů testových položek 3PL modelu, včetně parametru obtížnosti.*

### 3.3.3 Schopnost testové položky rozlišit mezi žáky podle jejich výsledků v ověřovacím testu, identifikace matoucí správné odpovědi a potenciálně chybný klíč hodnocení

Název modelové situace	
Schopnost testové položky rozlišit mezi žáky podle jejich výsledků v ověřovacím testu, identifikace matoucí správné odpovědi a potenciálně chybný klíč hodnocení	
Metodický postup řešení modelové situace	
Metodický postup řešení modelové situace je založen na výpočtu ukazatele schopnosti testové položky rozlišit mezi žáky s dobrými a slabými výsledky v ověřovacím testu, tj. na výpočtu ukazatele (parametru) diskriminace testové položky. Postup se skládá z následujících kroků uživatele metodiky:	
Krok 1: Uživatel metodiky vybírá, zda k výpočtu (odhadu) ukazatele diskriminace testové položky využije model vycházející z IRT.	
<i>Ne (přístup CTT)</i>	<i>Ano (přístup IRT)</i>
Krok 2: Uživatel metodiky počítá hodnoty ukazatelů diskriminace testové položky, tj.:	Krok 2: Uživatel metodiky odhaduje zvolený model v souladu s metodickým postupem modelové situace
(a) bodově biseriální korelace odpovědí žáků na testovou položku a výsledků žáků v celém testu,	„Volba a využití škály pro stanovení výsledků žáků v ověřovacím testu“.

<p>(b) upravené bodově biseriální korelace odpovědi žáků na testovou položku a výsledků žáků v celém testu s vynecháním vlastní hodnocené testové položky.</p>	<p>Krok 3: Z odhadu zvoleného modelu uživatel metodiky extrahuje parametr diskriminace testové položky.</p>
<b>Výstup řešení modelové situace</b>	
<p>Výstupem řešení modelové situace je hodnota bodově biseriální korelace či upravené bodově biseriální korelace jako ukazatelů schopnosti testových položek rozlišit mezi žáky s dobrými a slabými výsledky. Hodnota obou ukazatelů se pohybuje v intervalu od -1 do +1, přičemž vysoké kladné hodnoty naznačují dobrou schopnost diskriminace, zatímco záporné hodnoty nízkou schopnost diskriminace testové položky. Záporné hodnoty ukazují na silně problémovou testovou položku.</p>	<p>Výstupem řešení modelové situace je hodnota parametru diskriminace testové položky, která vyjadřuje, jak rychle se mění pravděpodobnost správné odpovědi na testovou položku (osa <math>y</math>) se změnou úrovně hodnoceného konstruktů (osa <math>x</math>). Obvyklé hodnoty parametru diskriminace se pohybují v intervalu hodnot od 0,5 do 2,5 (či 3,0) a platí, že vyšší hodnoty parametru diskriminace jsou spojeny s lepší schopností testové položky rozlišit mezi žáky s dobrými a slabými výsledky.</p>
<b>Širší situační kontext</b>	
<p>Ze dvou uvedených ukazatelů diskriminace testové položky založených na CTT je preferována upravená bodově biseriální korelace, která nezohledňuje vliv samotné testové položky na celkový výsledek v testu.</p> <p>Schopnost diskriminovat mezi žáky s dobrými a slabými výsledky v ověřovacím testu je jedním z klíčových parametrů charakterizujících kvalitu testové položky. Za málo přínosné jsou označovány testové položky s hodnotou bodově biseriální korelace nižší než 0,20 (přístup CTT). V případě přístupů založených na IRT je hodnota velmi nízké schopnosti testové položky diskriminovat mezi žáky s dobrými a slabými výsledky v ověřovacím testu uváděna jako nižší než 0,35.</p> <p>Schopnost testové položky diskriminovat mezi žáky s dobrými a slabými výsledky v ověřovacím testu ovlivňuje spolehlivost testu na různých úrovních zvládnutí hodnoceného konstruktů. Takto testové položky s vyššími hodnotami parametru diskriminace poskytují také více informací pro celý test, nejvíce pak na úrovni své obtížnosti. Podobně testové položky dobře diskriminující mezi žáky s dobrými a slabými výsledky v ověřovacím testu vykazují vyšší úroveň korelace s celkovými výsledky žáků, a zvyšují tak spolehlivost celého testu.</p> <p>Velmi nízké, či v případě ukazatelů založených na CTT dokonce záporné, hodnoty ukazatelů diskriminace testové položky naznačují možnou existenci problémů: (a) nevhodně formulované správné odpovědi testové položky; (b) silně matoucí nesprávné odpovědi (distraktory) testové položky; a (c) chybného klíče k hodnocení testové položky. V tomto případě je žádoucí další hodnocení testové položky: (a) prověření správnosti klíče hodnocení testové položky; a (b) hodnocení kvality nabízených nesprávných odpovědí (distraktorů) testové položky.</p>	

Software a ilustrace jeho využití pro řešení modelové situace	
R balíček <i>CTT</i> (viz Willse, 2018)	R balíček <i>ltm</i> (viz Rizopoulos, 2018)
<p><b>Přístup CTT pro datový rámec TACR_DATA</b></p> <pre>Diskriminace_CTT &lt;- itemAnalysis(TACR_DATA, itemReport=TRUE, rBisML = TRUE)</pre> <p><i># Funkce itemAnalysis (R balíček CTT) vede k výpočtu hodnot charakteristik kvality testových položek (včetně ukazatelů bodové a upravené bodové biseriální korelace), které jsou uloženy v objektu Diskriminace_CTT.</i></p> <p><i># Atribut itemReport uvádí, že v objektu Diskriminace_CTT má být uložen také report obsahující hodnoty charakteristik kvality testových položek.</i></p> <p><i># Atribut rBisML uvádí, že výpočet biseriálních korelací bude proveden s využitím metody maximální věrohodnosti, přičemž alternativou je rychlejší ad-hoc odhad.</i></p> <pre>Diskriminace_CTT\$itemReport\$pBis</pre> <p><i># Příkaz vede k „vytažení“ reportu obsahujícího hodnoty charakteristik kvality testových položek (itemReport) a z nich pak „vytažení“ hodnoty upravené bodové biseriální korelace (pBis).</i></p> <p><b>Přístup IRT pro datový rámec TACR_DATA a odhad 3PL modelu</b></p> <pre>PL3model &lt;- tpm(TACR_DATA)</pre> <p><i># Funkce tpm (R balíček ltm) vede k odhadu parametrů 3PL modelu, které jsou uloženy v objektu PL3model.</i></p> <pre>coef(PL3model)</pre> <p><i># Funkce coef (R balíček ltm) zobrazí hodnoty parametrů testových položek 3PL modelu, včetně parametru diskriminace.</i></p>	

### 3.3.4 Kvalita nabízených nesprávných odpovědí testové položky

Název modelové situace
Kvalita nabízených nesprávných odpovědí testové položky
Metodický postup řešení modelové situace
<p>Metodický postup řešení modelové situace je analogií k řešení modelové situace „Schopnost testové položky rozlišit mezi žáky podle jejich výsledků v ověřovacím testu, identifikace matoucí správné odpovědi a potenciálně chybný klíč hodnocení“. Rozdíl mezi oběma modelovými situacemi spočívá v tom, že jako „správná“ odpověď je v této modelové situaci označena nesprávná odpověď (distraktor), která je předmětem hodnocení. Dále uživatel metodiky sleduje následující kroky postupu:</p> <p>Krok 1: Uživatel metodiky určí podíl žáků účastnících se ověřovacího testu, kteří vybrali hodnocenou nesprávnou odpověď (distraktor) jako svou odpověď na danou testovou položku.</p> <p>Krok 2: Uživatel metodiky vybírá, zda bude kvalitu hodnocené nesprávné odpovědi (distraktoru) dané testové položky posuzovat s využitím odhadu modelu vycházejícího z IRT.</p>

<i>Ne (přístup CTT)</i>	<i>Ano (přístup IRT)</i>
<p>Krok 3: Uživatel metodiky počítá hodnoty ukazatelů diskriminace testové položky, tj.:</p> <p>(a) bodově biseriální korelace odpovědí žáků na testovou položku a výsledků žáků v celém testu,</p> <p>(b) upravené bodově biseriální korelace odpovědí žáků na testovou položku a výsledků žáků v celém testu s vynecháním vlastní hodnocené testové položky.</p>	<p>Krok 3: Uživatel metodiky odhaduje zvolený model v souladu s metodickým postupem modelové situace „Volba a využití škály pro stanovení výsledků žáků v ověřovacím testu“.</p> <p>Krok 4: Z odhadu zvoleného modelu uživatel metodiky extrahuje parametr diskriminace testové položky.</p>
<b>Výstup řešení modelové situace</b>	
<p>Výstupem řešení modelové situace je: (a) podíl žáků, kteří vybrali hodnocenou nesprávnou odpověď (distraktor) jako svou odpověď na danou testovou položku; a (b) hodnota bodově biseriální korelace či upravené bodově biseriální korelace testové položky s danou nesprávnou odpovědí (distraktorem). Kvalitu nesprávné odpovědi (distraktoru) zpochybňuje: (a) velmi nízký podíl žáků vybírajících nesprávnou odpověď (distraktor); a (b) vysoká kladná hodnota bodově biseriální korelace či upravené bodově biseriální korelace.</p>	<p>Výstupem řešení modelové situace je: (a) podíl žáků, kteří vybrali hodnocenou nesprávnou odpověď (distraktor) jako svou odpověď na danou testovou položku; a (b) hodnota parametru diskriminace testové položky s danou nesprávnou odpovědí (distraktorem). Kvalitu distraktoru zpochybňuje: (a) velmi nízký podíl žáků vybírajících nesprávnou odpověď (distraktor); a (b) vysoká kladná hodnota parametru diskriminace testové položky.</p>
<b>Širší situační kontext</b>	
<p>Ze dvou uvedených ukazatelů diskriminace testové položky založených na CTT je preferována upravená bodově biseriální korelace, která nezohledňuje vliv samotné testové položky na celkový výsledek v testu.</p> <p>Hodnocení nesprávných odpovědí (distraktorů) je žádoucí především u těch testových položek, které vykazují nízké, v případě CTT přístupů dokonce záporné, hodnoty ukazatelů diskriminace (upravená bodově biseriální korelace v případě přístupů CTT, parametr diskriminace v případě IRT přístupů). Detailní analýza nabízených odpovědí žákům může pomoci rozhodnout, zda takovou testovou položku ve vyhodnocení ověřovacího testu ponechat či nikoliv. Při přípravě ověřovacího testu, který by chtěl danou testovou položku využít, je žádoucí zvážit přeformulování nabízených odpovědí žákům.</p>	
<b>Software a ilustrace jeho využití pro řešení modelové situace</b>	
R balíček <i>CTT</i> (viz Willse, 2018)	R balíček <i>ltm</i> (viz Rizopoulos, 2018)
<p>Řešení modelové situace je analogií k řešení modelových situací v podkapitolách 3.3.2 a 3.3.3. Jediný rozdíl spočívá v kódování hodnocené testové položky, kdy správná odpověď má hodnotu „0“, zatímco posuzovaná nesprávná odpověď (distraktor) má hodnotu „1“.</p>	

Předpokládejme využití datového rámce TACR\_DATA\_DISTRAKT a hodnocení kvality nesprávné odpovědi A pro testovou položku item27 (proměnná na 39. místě datového rámce s označením item27A). Před využitím postupů uvedeného v podkapitolách 3.3.1 a 3.3.2 je vytvořen alternativní datový rámec TACR\_DATA\_27A:

```
TACR_DATA_27A <- TACR_DATA_DISTRAKT[-c(27,40,41,42,43,44)]
```

*# Příkaz vede k „vymazání“ proměnných v příslušném pořadí z datového rámce TACR\_DATA\_DISTRAKT a k vytvoření datového rámce TACR\_DATA\_27A, který obsahuje všechny testové položky základního testu s tím, že testová položka item27 má jako „správnou“ odpověď (hodnota „1“) označenu nesprávnou odpověď (distraktor) A.*

V dalším postupu je na všech relevantních místech nahrazen datový rámec TACR\_DATA za datový rámec TACR\_DATA\_27A.

### 3.3.5 Změna spolehlivosti ověřovacího testu při vynechání testové položky

Název modelové situace
Změna spolehlivosti ověřovacího testu při vynechání testové položky
Metodický postup řešení modelové situace
<p>Metodický postup řešení modelové situace je analogií modelové situace „Spolehlivost (škály) ověřovacího testu“, na kterou také logicky navazuje. Rozdíl obou modelových situací spočívá ve vynechání posuzované testové položky z hodnocení. Uživatel metodiky tak sleduje následující kroky postupu:</p> <p>Krok 1: Uživatel metodiky pracuje s ověřovacím testem, který obsahuje posuzovanou testovou položku. V souladu s metodickým postupem pro modelovou situaci „Spolehlivost (škály) ověřovacího testu“ uživatel metodiky vypočítá hodnoty šesti ukazatelů spolehlivosti testu (<math>\lambda_1</math> až <math>\lambda_6</math>), kde <math>\lambda_3</math> je Cronbachovo alfa, a dále dva ukazatele spolehlivosti testu <math>\omega_h</math> a <math>\omega_t</math>, kde se ukazatel <math>\omega_h</math> vztahuje pouze k hlavnímu konstruktů a ukazatel <math>\omega_t</math> zohledňuje také vliv skupinových konstruktů dílčích testových položek.</p> <p>Krok 2: Uživatel metodiky vynechá z ověřovacího testu posuzovanou testovou položku. Následně uživatel metodiky opakuje metodický postup kroku 1 pro ověřovací test, který posuzovanou testovou položku neobsahuje.</p> <p>Krok 3: Uživatel metodiky porovná hodnoty ukazatelů spolehlivosti testu (<math>\lambda_1</math> až <math>\lambda_6</math>), <math>\omega_h</math> a <math>\omega_t</math> získané z kroků 1 a 2 postupu, tj. pro ověřovací test obsahující a neobsahující posuzovanou testovou položku.</p>
Výstup řešení modelové situace
<p>Výstupem řešení modelové situace jsou dvě množiny ukazatelů spolehlivosti testu (<math>\lambda_1</math> až <math>\lambda_6</math>), <math>\omega_h</math> a <math>\omega_t</math>, a to pro:</p> <p>(a) ověřovací test, který obsahuje posuzovanou testovou položku;</p> <p>(b) ověřovací test, který neobsahuje posuzovanou testovou položku.</p> <p>Pokud jsou hodnoty ukazatelů spolehlivosti druhého z uvedených ověřovacích testů vyšší než korespondující hodnoty ověřovacího testu prvního, pak vynechání testové položky zvyšuje spolehlivost ověřovacího testu.</p>



Širší situační kontext
<p>Hodnocení změny spolehlivosti ověřovacího testu při vynechání dané testové položky dále doplňuje charakteristiky kvality testových položek. Takto se spolehlivost testu typicky zvyšuje při vynechání především těch testových položek, které mají nízkou schopnost diskriminovat mezi žáky v závislosti na jejich úrovni zvládnutí hodnoceného konstruktů. Za žádoucí lze považovat bližší zkoumání vlastností těch testových položek, jejichž vynechání zvyšuje spolehlivost ověřovacího testu.</p>
Software a ilustrace jeho využití pro řešení modelové situace
<p>R balíček <i>psych</i> (viz Revelle, 2020)</p>
<p>Řešení modelové situace je analogií k řešení modelové situace představené v podkapitole 3.3.1. Takto první krok metodického postupu řešení je stejný, jako je řešení modelové situace v podkapitole 3.3.1. Druhý krok metodického postupu řešení se pak odlišuje pouze ve vynechání posuzované testové položky z datového rámce TACR_DATA. Předpokládejme, že se bude jednat o testovou položku item27. V takovém případě je využít následující příkaz, který vytvoří nový datový rámec bez testové položky item27:</p> <pre>TACR_DATA_BEZ27 &lt;- TACR_DATA[-c(27)]</pre> <p><i># Příkaz vede k „vymazání“ proměnné item27 z datového rámce TACR_DATA a k vytvoření datového rámce TACR_DATA_BEZ27, který obsahuje všechny testové položky základního testu s výjimkou testové položky item27.</i></p> <p>Následně je zopakován metodický postup představený v podkapitole 3.1.1 pro nově uložený datový rámec TACR_DATA_BEZ27, který na všech relevantních místech nahradí datový rámec TACR_DATA.</p> <p>Získané množiny dat umožňují snadné srovnání hodnot podle třetího kroku metodického postupu řešení.</p>

### 3.3.6 (Ne)spravedlnost testových položek k různým skupinám žáků

Název modelové situace
(Ne)spravedlnost testových položek k různým skupinám žáků
Metodický postup řešení modelové situace
<p>Metodický postup řešení modelové situace je založen na analýze (DIF), která posuzuje vliv charakteristiky utvářející různé skupiny žáků (např. socioekonomický status rodiny) na správnost jejich odpovědí na hodnocenou testovou položku, přičemž kontrolován je vliv úrovně zvládnutí hodnoceného konstruktů žáky. Testová položka se nechová spravedlivě, pokud testovaní žáci z jedné skupiny a s danou úrovní zvládnutí hodnoceného konstruktů nemají stejnou pravděpodobnost správné odpovědi na hodnocenou testovou položku jako testovaní žáci z druhé skupiny a se stejnou úrovní zvládnutí hodnoceného konstruktů. Postup řešení modelové situace se pak skládá z následujících kroků uživatele metodiky:</p> <p>Krok 1: Uživatel metodiky zvolí tzv. ohniskovou skupinu žáků, u které předpokládá znevýhodnění a referenční skupinu žáků, vůči které je ohnisková skupina žáků posuzována.</p>

Krok 2: Uživatel metodiky vybírá, zda bude spravedlnost testové položky posuzovat s využitím odhadu modelu vycházejícího z IRT.

*Ne (přístup CTT)*

Krok 3: Uživatel metodiky volí metodu analýzy (DIF), kterou použije.

*Krok 3a): Metoda logistické regrese*

Krok 4a): Uživatel metodiky počítá úroveň DIF srovnáním hodnot pseudo- $R^2$  dvou regresních modelů: (a) modelu, který vysvětluje (log) pravděpodobnost správné odpovědi žáka na testovou položku jeho úrovní zvládnutí hodnoceného konstruktů (např. skóre v testu); a (b) modelu, který vysvětluje (log) pravděpodobnost správné odpovědi žáka na testovou položku jeho úrovní zvládnutí hodnoceného konstruktů (např. skóre v testu) a příslušností k ohniskové či referenční skupině, včetně interakčního členu dvou vysvětlujících proměnných.

Krok 5a): Uživatel metodiky posoudí úroveň DIF zařazením testové položky do jedné z kategorií:

- A – nízká úroveň pro hodnoty nižší než 0,035
- B – středně silná úroveň pro hodnoty 0,035 až 0,070
- C – vysoká úroveň pro hodnoty vyšší než 0,070

*Krok 3b): Mantel-Haenszelův (MH) přístup*

Krok 4b) Uživatel metodiky počítá hodnoty  $MH\chi^2$  statistiky a  $\Delta MH$  pro hodnocenou testovou položku.

Krok 5b) Uživatel metodiky klasifikuje testovou položku do jedné z kategorií:

- A – nízká úroveň DIF (statisticky nevýznamné hodnoty  $MH\chi^2$  a absolutní hodnota  $\Delta MH$  menší než 1)
- B – střední úroveň DIF (statisticky významné hodnoty  $MH\chi^2$  a absolutní hodnota  $\Delta MH$  menší než 1,5)
- C – vysoká úroveň DIF (statisticky významné hodnoty  $MH\chi^2$  a absolutní hodnota  $\Delta MH$  vyšší než 1,5)

*Ano (přístup IRT)*

Krok 3: Uživatel metodiky odhaduje zvolený model v souladu s metodickým postupem modelové situace „Volba a využití škály pro stanovení výsledků žáků v ověřovacím testu“.

Krok 4: Z odhadu zvoleného modelu uživatel metodiky extrahuje úroveň zvládnutí hodnoceného konstruktů žáky ( $\theta$ ).

Krok 5: Uživatel metodiky počítá úroveň DIF srovnáním hodnot pseudo- $R^2$  dvou regresních modelů: (a) modelu, který vysvětluje (log) pravděpodobnost správné odpovědi žáka na testovou položku jeho úrovní zvládnutí hodnoceného konstruktů ( $\theta$ ); a (b) modelu, který vysvětluje (log) pravděpodobnost správné odpovědi žáka na testovou položku jeho úrovní zvládnutí hodnoceného konstruktů ( $\theta$ ) a příslušností k ohniskové či referenční skupině, včetně interakčního členu dvou vysvětlujících proměnných.

Krok 6: Uživatel metodiky posoudí úroveň DIF zařazením testové položky do jedné z kategorií:

- A – nízká úroveň pro hodnoty nižší než 0,035
- B – středně silná úroveň pro hodnoty 0,035 až 0,070
- C – vysoká úroveň pro hodnoty vyšší než 0,070

Krok 7: Při identifikaci vysoké úrovně DIF testové položky může uživatel metodiky posoudit, zda se výsledek žáků v řešení testové položky liší pro všechny úrovně zvládnutí hodnoceného konstruktů (uniformní DIF), nebo jen pro některé z nich (neuniformní DIF). Za tímto účelem uživatel metodiky rozdělí žáky ohniskové i referenční skupiny do konečného počtu tříd podle zvoleného kvantilu jejich výsledků a pro tyto třídy dopočítá a posoudí rozdíly v jejich schopnosti řešit testovou položku.

<i>Společně pro oba metodické přístupy</i>	
<p>Krok 6: Při identifikaci vysoké úrovně DIF testové položky může uživatel metodiky posoudit, zda se výsledek žáků v řešení testové položky liší pro všechny úrovně zvládnutí hodnoceného konstrukturu (uniformní DIF), nebo jen pro některé z nich (neuniformní DIF). Za tímto účelem uživatel metodiky rozdělí žáky ohniskové i referenční skupiny do konečného počtu tříd podle zvoleného kvantilu jejich výsledků a pro tyto třídy dopočítá a posoudí rozdíly v jejich schopnosti řešit testovou položku.</p>	
<b>Výstup řešení modelové situace</b>	
<p>Výstupem řešení modelové situace je klasifikace testových položek podle úrovně DIF do kategorií A, B a C, a to při aplikaci metody logistické regrese i MH přístupu. Zařazení testové položky do kategorie C naznačuje vysokou úroveň DIF, což zpochybňuje spravedlnost testové položky. Lze doplnit informaci o uniformním či neuniformním DIF podle zjištění kroku 6 postupu.</p>	<p>Výstupem řešení modelové situace je klasifikace testových položek podle úrovně DIF do kategorií A, B a C. Zařazení testové položky do kategorie C naznačuje vysokou úroveň DIF, což zpochybňuje spravedlnost testové položky. Lze doplnit informaci o uniformním či neuniformním DIF podle zjištění kroku 6 postupu.</p>
<b>Širší situační kontext</b>	
<p>Modelová situace je zasazena do širšího rámce konceptu spravedlnosti ve vzdělávání, která je spojována s různými možnými zdroji nespravedlnosti – např. pohlaví žáků, jazykový původ, socioekonomické znevýhodnění. Tyto zdroje nespravedlnosti je žádoucí vzít do úvahy při hodnocení a tvorbě testových položek.</p> <p>Pokud vyšší počet testových položek znevýhodňuje určitou skupinu žáků, mohou tyto testové položky utvářet vedlejší konstrukt a narušovat předpoklad unidimenzionality testu.</p> <p>Spravedlnost testové položky je další významnou vlastností, která charakterizuje kvalitní testové položky. V případě identifikace vysoké úrovně DIF je žádoucí bližší analýza zdrojů nespravedlnosti a následně případné přeformulování znění testové položky.</p>	
<b>Software a ilustrace jeho využití pro řešení modelové situace</b>	
R balíček <i>difR</i> (viz Magis, Beland a Raiche, 2020)	R balíček <i>difR</i> (viz Magis, Beland a Raiche, 2020)
<p><b>Přístup CTT pro datový rámec TACR_DATA_DIF – logistická regrese</b></p> <pre>difLogistic(TACR_DATA_DIF, group = "divka", focal.name = 1, type = "both", alpha=0.01)</pre> <p><i># Funkce difLogistic (R balíček difR) vede k výpočtu hodnot následujících výstupů DIF analýzy pro všechny testové položky datového rámce TACR_DATA_DIF: (a) LTR statistika či Waldova statistika a související p-hodnoty pro posouzení</i></p>	

statistické významnosti obou statistik; a (b) klasifikace testových položek do kategorie A, B, C na základě srovnání Cox-Snellova  $R^2$  dvou odhadovaných modelů (A – 0 až 0,034; B – 0,035 až 0,069; C – 0,070 a více).

# Atribut **group** definuje proměnnou (divka), vůči které je DIF analýza testových položek prováděna.

# Atribut **focal.name** definuje ohniskovou skupinu, kterou jsou dívky („1“), referenční skupinou jsou chlapci („0“).

# Atribut **type** určuje, že hodnocení bere do úvahy uniformní i neuniformní DIF („both“) s tím, že alternativami jsou: (a) pouze uniformní DIF (type = "udif"); a (b) pouze neuniformní DIF (type = "nudif").

# Atribut **alpha** určuje hladinu (0,01) pro rozhodnutí o statistické významnosti LRT statistiky a Waldovy statistiky.

#### **Přístup CTT pro datový rámec TACR\_DATA\_DIF – Mantel-Haenszelův (MH) přístup**

difMH(TACR\_DATA\_DIF, group = "divka", focal.name = 1, alpha=0.01)

# Funkce difMH (R balíček *difR*) vede k výpočtu hodnot následujících výstupů DIF analýzy pro všechny testové položky datového rámce TACR\_DATA\_DIF: (a)  $MH_{\chi^2}$  statistika a p-hodnota k ní pro posouzení statistické významnosti; a (b) klasifikace testových položek do kategorie A, B, C v závislosti na statistické významnosti  $MH_{\chi^2}$  statistiky a na hodnotě  $\Delta MH$ .

# Atribut **group** definuje proměnnou (divka), vůči které je DIF analýza testových položek prováděna.

# Atribut **focal.name** definuje ohniskovou skupinu, kterou jsou dívky („1“), referenční skupinou jsou chlapci („0“).

# Atribut **alpha** určuje hladinu (0,01) pro rozhodnutí o statistické významnosti  $MH_{\chi^2}$  statistiky.

Postup řešení s využitím modelu, který vychází z IRT, je uveden v kartě modelové situace. Příklad ilustrující toto řešení s využitím programovacího jazyka R zde není uveden. K řešení je však možné využít celou řadu R balíčků zaměřených na odhady logistických regresních modelů.

### **3.3.7 Kvalita testových položek a identifikace nekvalitních testových položek**

Název modelové situace
Kvalita testových položek a identifikace nekvalitních testových položek
Metodický postup řešení modelové situace
Metodický postup řešení modelové situace je založen na výpočtu hodnot ukazatelů (parametrů), které charakterizují kvalitu testových položek: (a) obtížnost testových položek; (b) schopnost testových položek rozlišit mezi žáky s dobrými a slabými výsledky v ověřovacím testu; (c) kvalita nabízených nesprávných odpovědí (distraktorů) na testové položky; (d) změna spolehlivosti ověřovacího testu při vynechání testových položek; a (e) spravedlnost testových položek k různým skupinám žáků. V případě aplikace přístupů vycházejících z IRT je rovněž kvalita testové položky hodnocena prostřednictvím statistik (indexů) dobré shody skutečných a modelem predikovaných odpovědí žáků na testovou položku.
Při výpočtu hodnot ukazatelů pro charakteristiky kvality testových položek (a) až (e) jsou aplikovány metodické postupy pěti souvisejících modelových situací: (a) „Zvládnutí testové položky žáky a stanovení obtížnosti

testové položky“; (b) „Schopnost testové položky rozlišit mezi žáky podle jejich výsledků v ověřovacím testu, identifikace matoucí správné odpovědi a potenciálně chybný klíč hodnocení“; (c) „Kvalita nabízených nesprávných odpovědí testové položky“; (d) „Změna spolehlivosti ověřovacího testu při vynechání testové položky“; a (e) „Hodnocení spravedlnosti testových položek k charakteristikám různých skupin žáků“. Takto při řešení modelové situace sleduje uživatel metodiky následující kroky postupu:

Krok 1: Uživatel metodiky vybírá, zda bude kvalitu testových položek posuzovat s využitím odhadu modelu vycházejícího z IRT.

<i>Ne (přístup CTT)</i>	<i>Ano (přístup IRT)</i>
<p>Krok 2: Uživatel metodiky aplikuje metodický postup řešení modelové situace „Zvládnutí testové položky žáky a stanovení obtížnosti testové položky“ pro přístup CTT.</p>	<p>Krok 2: Uživatel metodiky aplikuje metodický postup řešení modelové situace „Zvládnutí testové položky žáky a stanovení obtížnosti testové položky“ pro přístup IRT.</p>
<p>Krok 3: Uživatel metodiky aplikuje metodický postup řešení modelové situace „Schopnost testové položky rozlišit mezi žáky podle jejich výsledků v ověřovacím testu, identifikace matoucí správné odpovědi a potenciálně chybný klíč hodnocení“ pro přístup CTT.</p>	<p>Krok 3: Uživatel metodiky aplikuje metodický postup řešení modelové situace „Schopnost testové položky rozlišit mezi žáky podle jejich výsledků v ověřovacím testu, identifikace matoucí správné odpovědi a potenciálně chybný klíč hodnocení“ pro přístup IRT.</p>
<p>Krok 4: Uživatel metodiky aplikuje metodický postup řešení modelové situace „Kvalita nabízených nesprávných odpovědí testové položky“ pro přístup CTT.</p>	<p>Krok 4: Uživatel metodiky aplikuje metodický postup řešení modelové situace „Kvalita nabízených nesprávných odpovědí testové položky“ pro přístup IRT.</p>
<p>Krok 5: Uživatel metodiky aplikuje metodický postup řešení modelové situace „Změna spolehlivosti ověřovacího testu při vynechání testové položky“ pro přístup CTT.</p>	<p>Krok 5: Uživatel metodiky aplikuje metodický postup řešení modelové situace „Hodnocení spravedlnosti testových položek k charakteristikám různých skupin žáků“ pro přístup IRT.</p>
<p>Krok 6: Uživatel metodiky aplikuje metodický postup řešení modelové situace „Hodnocení spravedlnosti testových položek k charakteristikám různých skupin žáků“ pro přístup CTT.</p>	<p>Krok 6: Uživatel metodiky posuzuje úroveň dobré shody skutečných a zvoleným modelem predikovaných odpovědí žáků na testovou položku výpočtem odhadů tří statistik: (a) Yenova <math>Q_1</math>; (b) <math>S-\chi^2</math> statistiky; a (c) PV-<math>Q_1</math> statistiky. K těmto statistikám uživatel metodiky dále stanovuje jejich statistickou významnost a hodnotu RMSEA.</p>
<b>Výstup řešení modelové situace</b>	
<p>Výstupem řešení modelové situace je charakteristika testových položek prostřednictvím:</p>	<p>Výstupem řešení modelové situace je charakteristika testových položek prostřednictvím:</p> <p>(a) hodnot ukazatele obtížnosti testových položek vyjádřeného jako úroveň zvládnutí hodnoceného</p>

<p>(a) hodnot ukazatele obtížnosti testových položek vyjádřeného jako podíl žáků, kteří zodpověděli danou testovou položku správně;</p> <p>(b) hodnot ukazatele diskriminace testových položek vyjádřeného jako hodnota bodově biseriální korelace či upravené bodově biseriální korelace;</p> <p>(c) identifikace nabízených nesprávných odpovědí (distraktorů) na testové položky, které vykazují nepříznivé vlastnosti;</p> <p>(d) hodnot změny ukazatelů spolehlivosti testu při vynechání testových položek;</p> <p>(e) zařazení testových položek do kategorií A, B a C vzhledem k úrovni DIF.</p>	<p>konstruktů, na které přibližně 50 % testovaných žáků odpovídá testovou položku správně;</p> <p>(b) hodnot ukazatele diskriminace testových položek vyjadřujícího rychlost změny pravděpodobnosti správné odpovědi žáka se změnou jeho úrovně zvládnutí hodnoceného konstruktů;</p> <p>(c) identifikace nabízených nesprávných odpovědí (distraktorů) na testové položky, které vykazují nepříznivé vlastnosti;</p> <p>(d) zařazení testových položek do kategorií A, B a C vzhledem k úrovni DIF;</p> <p>(e) statistické významnosti a hodnoty RMSEA statistik Yenova <math>Q_1</math>, <math>S-\chi^2</math> a <math>PV-Q_1</math>.</p>
--	---

#### Širší situační kontext

Řešení modelové situace poskytuje komplexní pohled na kvalitu testových položek ověřovacího testu a umožňuje identifikovat testové položky, jejichž kvalitu je nutné hodnotit jako horší. V tomto ohledu platí:

- V případě využití přístupů založených na CTT uživatel metodiky věnuje pozornost kvalitě testových položek charakteristických: (a) hodnotou ukazatele obtížnosti vyšší než 0,90 a nižší než 0,10 (0,20); (b) hodnotou ukazatele diskriminace nižší než 0,20, speciálně pak zápornou hodnotou ukazatele diskriminace; (c) přítomností nabízených nesprávných odpovědí (distraktorů) s nepříznivými vlastnostmi; (d) zvýšením spolehlivosti testu při jejich vynechání; a (e) zařazením do C kategorie vzhledem k úrovni DIF.
- V případě využití přístupů založených na IRT uživatel metodiky věnuje pozornost kvalitě testových položek charakteristických: (a) hodnotou ukazatele obtížnosti vyšší než +2 (+3) a nižší než -2 (-3); (b) hodnotou ukazatele diskriminace nižší než 0,35; (c) přítomností nabízených nesprávných odpovědí (distraktorů) s nepříznivými vlastnostmi; (d) zařazením do C kategorie vzhledem k úrovni DIF; a (e) hodnotou RMSEA statistik Yenova  $Q_1$ ,  $S-\chi^2$  a  $PV-Q_1$  vyšší než 0,05.

Uživatel metodiky využívá uvedené informace při vyhodnocení ověřovacího testu, kdy především rozhoduje o způsobu vyhodnocení nekvalitních testových položek, a dále pak při utváření nových testů, kdy mu informace o kvalitě testových položek umožňují rozhodovat o nejhodnější podobě ověřovacího testu. Takto může být například záměr cílit obtížnost ověřovacího testu na stanovenou úroveň zvládnutí hodnoceného konstruktů reflektován zařazením vyššího počtu testových položek s odpovídající obtížností.

Analogicky k vykreslení informační křivky testu lze vykreslit i informační křivku testových položek (viz modelová situace „Spolehlivost (škály) ověřovacího testu“).

Software a ilustrace jeho využití pro řešení modelové situace	
R balíček CTT (viz Willse, 2018)	R balíček ltm (viz Rizopoulos, 2018)
R balíček psych (viz Revelle, 2020)	R balíček difR (viz Magis, Beland a Raiche, 2020)
R balíček difR (viz Magis, Beland a Raiche, 2020)	R balíček mirt (Chalmers, 2020)
<p>Řešení modelové situace je shodné s řešením modelových situací v podkapitolách 3.3.1 až 3.3.6. Jedinou výjimkou je hodnocení úrovně dobré shody skutečných a zvoleným modelem predikovaných odpovědí žáků na testovou položku.</p> <p><b>Přístup IRT pro hodnocení úrovně dobré shody skutečných a zvoleným modelem predikovaných odpovědí žáků na testovou položku pro datový rámec TACR_DATA</b></p> <pre>PL2model &lt;- mirt(TACR_DATA, 1)  # Funkce mirt (R balíček mirt) vede k odhadu parametrů 2PL modelu, které jsou uloženy v objektu PL2model.  # Atribut 1 vyjadřuje, že má být odhadován unidimenzionální 2PL model.  itemfit(PL2model, fit_stats = "X2")  # Funkce itemfit (R balíček mirt) s atributem fits_stats = "X2" vede k zobrazení hodnot statistiky Yenova Q1, a to včetně p-hodnoty této statistiky a hodnoty RMSEA.  itemfit(PL2model, fit_stats = "S_X2")  # Funkce itemfit (R balíček mirt) s atributem fits_stats = " S_X2" vede k zobrazení hodnot statistiky S-χ², a to včetně p-hodnoty této statistiky a hodnoty RMSEA.  itemfit(PL2model, fit_stats = "PV_Q1")  # Funkce itemfit (R balíček mirt) s atributem fits_stats = "PV_Q1" vede k zobrazení hodnot statistiky PV-Q1, a to včetně p-hodnoty této statistiky a hodnoty RMSEA.</pre>	

### 3.3.8 Neobvyklý vzor odpovědí žáků na testové položky

Název modelové situace
Neobvyklý vzor odpovědí žáků na testové položky
Metodický postup řešení modelové situace
Metodický postup řešení modelové situace vychází z hledání neobvyklého vzoru odpovědí žáků na testové položky ověřovacího testu. Neobvyklý vzor odpovědí žáků je charakteristický tím, že žák zodpovídá obtížné testové položky správně, zatímco jednoduché testové položky chybně. Postup řešení modelové situace se pak skládá z následujících kroků uživatele metodiky:



Krok 1: Uživatel metodiky vybírá, zda bude neobvyklý vzor odpovědí žáků hodnotit s využitím odhadu modelu vycházejícího z IRT.	
<i>Ne (přístup CTT)</i>	<i>Ano (přístup IRT)</i>
<p>Krok 2: Uživatel metodiky posuzuje neobvyklý vzor odpovědí žáka prostřednictvím výpočtu:</p> <p>(a) bodově biseriální korelace mezi odpověďmi žáka na testové položky a hodnotami obtížnosti testových položek (statistika <math>r_{pbis}</math>),</p> <p>(b) hodnot statistik <math>C</math>, <math>C^*</math> a <math>U3</math>, které posuzují soulad mezi vzorem odpovědí žáků a tzv. Guttmanovým vzorem odpovědí, který odpovídá ideálnímu vzoru odpovědí žáků za předpokladu, že nejvíce obtížné testové položky zodpovídá žák chybně a naopak nejméně obtížné testové položky správně. Počet správných odpovědí žáka odpovídá jeho skóre v testu.</p>	<p>Krok 2: Uživatel metodiky odhaduje zvolený model v souladu s metodickým postupem modelové situace „Volba a využití škály pro stanovení výsledků žáků v ověřovacím testu“.</p> <p>Krok 3: Uživatel metodiky počítá hodnotu statistiky <math>Iz^*</math>, která posuzuje shodu mezi skutečnými odpověďmi žáků na testové položky a modelem predikovanými odpověďmi žáků na testové položky.</p>
<b>Výstup řešení modelové situace</b>	
Výstupem řešení modelové situace jsou hodnoty statistik $r_{pbis}$ , $C$ , $C^*$ a $U3$ pro žáky. Platí, že neobvyklý vzor odpovědí žáků je: (a) v případě statistiky $r_{pbis}$ spojený s vysokými zápornými hodnotami; a (b) v případě statistik $C$ , $C^*$ a $U3$ spojený s hodnotami, které se nejvíce odlišují od ideální hodnoty 0.	Výstupem řešení modelové situace jsou hodnoty statistiky $Iz^*$ pro žáky. Platí, že neobvyklý vzor odpovědí žáků je spojený s nízkými hodnotami statistiky $Iz^*$ .
<b>Širší situační kontext</b>	
<p>Modelová situace umožňuje identifikovat žáky, jejichž vzor odpovědí na testové položky ověřovacího testu je neobvyklý až podivný. Vynechání těchto žáků z ověřovacího testu vede ke zvýšení jeho spolehlivosti, takovému kroku však přirozeně předchází posouzení relevance takového rozhodnutí.</p> <p>Při identifikaci neobvyklého vzoru odpovědí žáků je potřeba vzít do úvahy také počet jimi správně zodpovězených otázek. Uvedené statistiky se mohou chovat různě například v případě žáků, jejichž úspěšnost v ověřovacím testu byla velmi nízká. Vhodné tak je posuzovat neobvyklý vzor odpovědí žáka na testové s využitím vyššího počtu statistik, respektive s využitím dalších podpůrných informací, které jsou k dispozici (např. doba řešení testu).</p>	
<b>Software a ilustrace jeho využití pro řešení modelové situace</b>	
R balíček <i>PerFit</i> (viz Tendeiro, 2018)	R balíček <i>PerFit</i> (viz Tendeiro, 2018)



**Přístup CTT pro datový rámec TACR\_DATA**

```
RPBIS <- r.pbis(TACR_DATA)
```

*# Funkce r.pbis (R balíček PerFit) vede k výpočtu hodnot bodově biseriální korelace (r.pbis) všech žáků. Hodnoty jsou uloženy v objektu RPBIS.*

```
write.csv(RPBIS$PFscores, "TACR_DATA_rpbis.csv")
```

*# Příkaz vede k „vytažení“ hodnot bodově biseriální korelace (r.pbis) všech žáků (PFscores) z datového rámce RPBIS s jejich uložením v csv souboru TACR\_DATA\_rpbis.*

```
C <- C.Sato(TACR_DATA)
```

*# Funkce C.Sato (R balíček PerFit) vede k výpočtu hodnot statistiky C všech žáků. Hodnoty jsou uloženy v objektu C.*

```
write.csv(C$PFscores, "TACR_DATA_C.csv")
```

*# Příkaz vede k „vytažení“ hodnot statistiky C všech žáků (PFscores) z datového rámce C s jejich uložením v csv souboru TACR\_DATA\_C.*

```
CSTAR <- Cstar(TACR_DATA)
```

*# Funkce Cstar (R balíček PerFit) vede k výpočtu hodnot statistiky C\* všech žáků. Hodnoty jsou uloženy v objektu CSTAR.*

```
write.csv(CSTAR$PFscores, "TACR_DATA_cstar.csv")
```

*# Příkaz vede k „vytažení“ hodnot statistiky C\* všech žáků (PFscores) z datového rámce CSTAR s jejich uložením v csv souboru TACR\_DATA\_cstar*

```
U3 <- U3(TACR_DATA)
```

*# Funkce U3 (R balíček PerFit) vede k výpočtu hodnot statistiky U3 všech žáků. Hodnoty jsou uloženy v objektu U3.*

```
write.csv(U3$PFscores, "TACR_DATA_U3.csv")
```

*# Příkaz vede k „vytažení“ hodnoty statistiky U3 všech žáků (PFscores) z datového rámce U3 s jejich uložením v csv souboru TACR\_DATA\_U3.*

**Přístup IRT pro datový rámec TACR\_DATA a odhad 2PL modelu**

```
PL2_LZSTAR <- lzstar(TACR_DATA, IRT.PModel = "2PL")
```

*# Funkce lzstar (R balíček PerFit) vede k výpočtu hodnot statistiky lz\* všech žáků. Hodnoty jsou uloženy v objektu PLS2\_LZSTAR.*

*# Atribut IRT.PMODEL specifikuje, který model vycházející z IRT bude odhadován. Vedle odhadovaného 2PL modelu jsou další možnosti 1PL model (IRT.PModel = "1PL") a 3PL model (IRT.PModel = "3PL")*

```
write.csv(PL2_LZSTAR$PFscores, "TACR_DATA_lzstar.csv")
```

*# Příkaz vede k „vytažení“ hodnoty statistiky lz\* všech žáků (PFscores) z datového rámce PLS2\_LZSTAR s jejich uložením v csv souboru TACR\_DATA\_lzstar.*

## 3.3.9 Volba a využití škály pro stanovení výsledků žáků v ověřovacím testu

Název modelové situace	
Volba a využití škály pro stanovení výsledků žáků v ověřovacím testu	
Metodický postup řešení modelové situace	
Metodický postup řešení modelové situace vychází z možnosti využití různých škál pro posouzení úrovně zvládnutí hodnoceného konstruktů žáky, tj. pro stanovení výsledků žáků v ověřovacím testu. Postup řešení modelové situace se skládá z následujících kroků uživatele metodiky:	
Krok 1: Uživatel metodiky vybírá, zda bude hodnotit výsledky žáků na škále s využitím odhadu modelu vycházejícího z IRT.	
<i>Ne (přístup CTT)</i>	<i>Ano (přístup IRT)</i>
Krok 2: Uživatel metodiky volí podobu škály:	Krok 2: Uživatel metodiky ověřuje předpoklady pro vyhodnocení ověřovacího testu s využitím modelů vycházejících z IRT, především pak:
<i>Krok 2a): Škála odpovídající dosaženému skóre žáka v testu</i>	vyhodnocení ověřovacího testu s využitím modelů vycházejících z IRT, především pak:
Krok 3a): Uživatel metodiky přiřadí každému žákovi jeho dosažené skóre v testu, tj. počet správně zodpovězených testových položek.	(a) předpoklad minimální požadované velikosti výběrového souboru žáků (viz kapitola 3.1);
<i>Krok 2b): Škála odpovídající procentuální úspěšnosti žáka v testu</i>	(b) předpoklad unidimenzionality testu hodnocený metodickým postupem modelové situace „Unidimenzionalita ověřovacího testu a počet konstruktů v něm obsažených“;
Krok 3b): Uživatel metodiky přiřadí každému žákovi podíl správných odpovědí, kterých dosáhl v řešení testu s vyjádřením v %.	(c) předpoklad lokální nezávislosti testových položek hodnocený metodickým postupem modelové situace „Lokální nezávislost testových položek“.
<i>Společně pro oba metodické přístupy</i>	<i>Krok 3a): Uživatel metodiky rozhodl o splnění předpokladů uvedených v kroku 2 nebo přijímá opatření pro naplnění předpokladů pro odhad modelů v souladu s předpoklady kroku 2.</i>
Krok 4: Uživatel metodiky může transformovat škálu kroků 2a) a 3a), respektive 2b) a 3b) na:	Krok 4a): Uživatel metodiky odhaduje parametry:
(a) alternativní bodovou škálu s průměrnou hodnotou $n$ bodů a směrodatnou odchylkou $m$ bodů;	(a) 1PL modelu; (b) 2PL modelu; (c) 3PL modelu.
(b) škálu percentilového pořadí.	Krok 5a): Uživatel metodiky vybírá nejvhodnější model v souladu s metodickým postupem modelové situace „Výběr nejvhodnějšího modelu vycházejícího z IRT a hodnocení úrovně dobré shody dat modelu a ověřovacího testu“.
Pozn.: V případě přístupu CTT lze, podobně jako při využití přístupu IRT, reagovat na poznatky z posouzení narušení předpokladů pro vyhodnocení ověřovacího testu, především předpokladu unidimenzionality.	

	<p>Krok 6a): Uživatel metodiky extrahuje hodnoty úrovně zvládnutí hodnoceného konstruktů žáky (<math>\Theta</math>) z odhadů parametrů modelu vybraného v kroku 5a).</p> <hr/> <p><i>Krok 3b): Uživatel metodiky rozhodl o nesplnění předpokladů uvedených v kroku 2 a odhaduje multidimenzionální model vycházející z IRT.</i></p> <hr/> <p>Krok 4b) Uživatel metodiky odhaduje parametry multidimenzionálního modelu vycházejícího z paradigmatu IRT, a to pro počet faktorů stanovených v souladu s metodickým postupem modelové situace „Unidimenzionalita ověřovacího testu a počet konstruktů v něm obsažených“.</p> <p>Krok 5b) Uživatel metodiky extrahuje hodnoty úrovně zvládnutí hodnocených konstruktů žáky (<math>\Theta_i</math>) z multidimenzionálního modelu odhadovaného v kroku 4b).</p> <hr/> <p><i>Společně pro metodické přístupy (a) a (b)</i></p> <hr/> <p>Krok 7a) / 6b): Uživatel metodiky může transformovat škály kroků 6a) a 5b) na:</p> <p>(a) alternativní bodovou škálu s průměrnou hodnotou <math>n</math> bodů a směrodatnou odchylkou <math>m</math> bodů;</p> <p>(b) škálu percentilového pořadí.</p> <hr/> <p><i>Krok 3c): Uživatel metodiky rozhodl o nesplnění předpokladů uvedených v kroku 2 a hodnotí výsledky na škále s využitím přístupu CTT.</i></p>
<b>Výstup řešení modelové situace</b>	
<p>Výstupem řešení modelové situace jsou hodnoty odpovídající úrovni zvládnutí hodnoceného konstruktů žáky v podobě: (a) skóre; (b) procentuální úspěšnost; (c) počet bodů na alternativní bodové škále; (d) percentilové umístění.</p>	<p>Výstupem řešení modelové situace jsou hodnoty odpovídající úrovni zvládnutí hodnoceného konstruktů žáky (<math>\Theta</math>), s případným vyjádřením v podobě: (a) počtu bodů na alternativní bodové škále; a (b) percentilového umístění.</p>

## Širší situační kontext

Při volbě škály pro stanovení výsledků žáků v ověřovacím testu je žádoucí posoudit naplnění předpokladů správnosti odhadů s danou škálou souvisejících (např. předpoklad unidimenzionality).

Při volbě nevhodnější škály vycházející z odhadů modelů založených na IRT lze pro rozhodnutí využít postup hodnocení dobré shody empirických dat a dat modelových.

Využití škál vycházejících z odhadů multidimenzionálních modelů založených na IRT má úzkou vazbu na hodnocení optimálního počtu konstruktů (dimenzí, faktorů) v ověřovacím testu obsažených.

Volba škály pro stanovení výsledků žáků v ověřovacím testu, tj. úrovně zvládnutí hodnoceného konstruktů, může být spojena s odlišnými závěry, a to především při využití odhadů komplexnějších modelů, které například zohledňují schopnost testových položek diferencovat mezi žáky podle úrovně zvládnutí hodnoceného konstruktů (např. 2PL a 3PL model). Takto žák dosahující lepšího výsledku na škále procentuální úspěšnosti řešení testových položek testu může dosáhnout horšího výsledku na škále vycházející z odhadu 2PL modelu. Tato skutečnost se může projevit také v dalších modelových situacích, v nichž se pracuje s výsledky žáků, jako je například záměr identifikovat faktory, které ovlivňují úrovně zvládnutí hodnoceného konstruktů žáky.

Moderní přístupy k reportingu výsledků žáků v ověřovacím testování se snaží spíše vyhýbat škálám založeným na skóre žáků či procentuální úspěšnosti.

## Software a ilustrace jeho využití pro řešení modelové situace

R balíček *CTT* (viz Willse, 2018)

R balíček *ltm* (viz Rizopoulos, 2018)

R balíček *CTT* (viz Willse, 2018)

R balíček *mirt* (viz Chalmers, 2020)

Výsledky žáků v ověřovacím testu na škálách vycházejících z přístupů CTT lze snadno stanovit s využitím standardního software. Z tohoto důvodu se ilustrace řešení modelové situace zaměřuje na přístupy IRT.

**Přístup IRT pro datový rámec TACR\_DATA – 1PL model**

```
PL1model <- rasch(TACR_DATA)
```

*# Funkce rasch (R balíček ltm) vede k odhadu parametrů 1PL modelu, které jsou uloženy v objektu PL1model.*

```
PL1model_vzor_skore <- factor.scores(PL1model, method = "EAP")
```

*# Funkce factor.scores (R balíček ltm) vede k odhadu úrovně zvládnutí hodnoceného konstruktů ( $\theta$ ) ve vazbě na vzor odpovědí na testové položky, přičemž využita jsou data z objektu PL1model odhadovaného v předchozím kroku metodického postupu. Výsledky jsou uloženy v objektu PL1model\_vzor\_skore.*

*# Argument method umožňuje volit metodu odhadu úrovně zvládnutí hodnoceného konstruktů ( $\theta$ ), v tomto případě se jedná o odhad založený na střední hodnotě aposteriorního rozdělení (EAP). Pro přístup založený na Bayesovském modálním odhadu (MAP) je volena možnost method = "EB".*

```
PL1theta <- PL1model_vzor_skore$score.dat
```

*# Příkaz vede k „vytažení“ hodnot úrovně zvládnutí hodnoceného konstruktů ( $\Theta$ ) pro všechny vzory odpovědí na testové položky a jejich uložení v objektu PL1theta.*

```
write.csv(PL1theta, "score_PL1_EAP.csv")
```

*# Příkaz vede k uložení hodnot úrovně zvládnutí hodnoceného konstruktů ( $\Theta$ ) pro všechny vzory odpovědí na testové položky a k jejich uložení v csv souboru score\_PL1\_EAP.*

Data souboru score\_PL1\_EAP.csv, která spojují vzory odpovědí s příslušnou úrovní zvládnutí hodnoceného konstruktů ( $\Theta$ ), jsou využita pro přiřazení hodnoty zvládnutí hodnoceného konstruktů každému žákovi.

### **Přístup IRT pro datový rámec TACR\_DATA – 2PL model**

```
PL2model <- ltm(TACR_DATA ~ z1)
```

*# Funkce ltm (R balíček ltm) vede k odhadu parametrů 2PL modelu, které jsou uloženy v objektu PL2model.*

*# Parametr z1 odpovídá úrovni zvládnutí hodnoceného konstruktů.*

```
PL2model_vzor_skore <- factor.scores(PL2model, method = "EAP")
```

*# Funkce factor.scores (R balíček ltm) vede k odhadu úrovně zvládnutí hodnoceného konstruktů ( $\Theta$ ) ve vazbě na vzor odpovědí na testové položky, přičemž využita jsou data z objektu PL2model odhadovaného v předchozím kroku metodického postupu. Výsledky jsou uloženy v objektu PL2model\_vzor\_skore.*

*# Atribut method umožňuje volit metodu odhadu úrovně zvládnutí hodnoceného konstruktů ( $\Theta$ ), v tomto případě se jedná o odhad založený na střední hodnotě aposteriorního rozdělení (EAP). Pro přístup založený na Bayesovském modálním odhadu (MAP) je volena možnost method = "EB".*

```
PL2theta <- PL2model_vzor_skore$score.dat
```

*# Příkaz vede k „vytažení“ hodnot úrovně zvládnutí hodnoceného konstruktů ( $\Theta$ ) pro všechny vzory odpovědí na testové položky a jejich uložení v objektu PL2theta.*

```
write.csv(PL2theta, "score_PL2_EAP.csv")
```

*# Příkaz vede k uložení hodnot úrovně zvládnutí hodnoceného konstruktů ( $\Theta$ ) pro všechny vzory odpovědí na testové položky a k jejich uložení v csv souboru score\_PL2\_EAP.*

Data souboru score\_PL2\_EAP.csv, která spojují vzory odpovědí s příslušnou úrovní zvládnutí hodnoceného konstruktů ( $\Theta$ ), jsou využita pro přiřazení hodnoty zvládnutí hodnoceného konstruktů každému žákovi.

### **Přístup IRT pro datový rámec TACR\_DATA – 3PL model**

```
PL3model <- tpm(TACR_DATA)
```

*# Funkce tpm (R balíček ltm) vede k odhadu parametrů 3PL modelu, které jsou uloženy v objektu PL3model.*

```
PL3model_vzor_skore <- factor.scores(PL3model, method = "EAP")
```

*# Funkce factor.scores (R balíček ltm) vede k odhadu úrovně zvládnutí hodnoceného konstruktů ( $\theta$ ) ve vazbě na vzor odpovědí na testové položky, přičemž využita jsou data z objektu PL3model odhadovaného v předchozím kroku metodického postupu. Výsledky jsou uloženy v objektu PL3model\_vzor\_skore.*

*# Atribut method umožňuje volit metodu odhadu úrovně zvládnutí hodnoceného konstruktů ( $\theta$ ), v tomto případě se jedná o odhad založený na střední hodnotě aposteriorního rozdělení (EAP). Pro přístup založený na Bayesovském modálním odhadu (MAP) je volena možnost method = "EB".*

```
PL3theta <- PL3model_vzor_skore$score.dat
```

*# Příkaz vede k „vytažení“ hodnot úrovně zvládnutí hodnoceného konstruktů ( $\theta$ ) pro všechny vzory odpovědí na testové položky a jejich uložení v objektu PL3theta.*

```
write.csv(PL3theta, "score_PL3_EAP.csv")
```

*# Příkaz vede k uložení hodnot úrovně zvládnutí hodnoceného konstruktů ( $\theta$ ) pro všechny vzory odpovědí na testové položky a k jejich uložení v csv souboru score\_PL3\_EAP.*

Data souboru score\_PL3\_EAP.csv, která spojují vzory odpovědí s příslušnou úrovní zvládnutí hodnoceného konstruktů ( $\theta$ ), jsou využita pro přiřazení hodnoty zvládnutí hodnoceného konstruktů každému žákovi.

#### **Přístup IRT pro datový rámec TACR\_DATA – multidimenzionální model**

```
MULTIDIMmodel <- mirt(TACR_DATA, 2, itemtype = "2PL")
```

*# Funkce mirt (R balíček mirt) vede k odhadu multidimenzionálního modelu vycházejícího z paradigmatu IRT, přičemž číslovka uvádí, kolika dimenzionální model má být odhadován, tj. zde 2 dimenze (konstrukty, faktory).*

*# Atribut itemtype umožňuje volit odhadovaný model založený na IRT, v tomto případě se jedná o 2PL model.*

```
summary(MULTIDIMmodel)
```

*# Obecná funkce summary vede k zobrazení faktorových zátěží testových položek k dimenzím (konstruktům, faktorům) multidimenzionálního modelu.*

```
theta <- fscores(MULTIDIMmodel, method = "EAP")
```

*# Funkce fscores (R balíček mirt) vede k odhadům úrovně zvládnutí hodnocených konstruktů žáky ( $\theta$ ), přičemž využita jsou data z objektu MULTIDIMmodel odhadovaného v předchozím kroku metodického postupu. Výsledky jsou uloženy v objektu theta.*

*# Atribut method umožňuje volit metodu odhadu úrovně zvládnutí hodnoceného konstruktů ( $\theta$ ), v tomto případě se jedná o odhad založený na střední hodnotě aposteriorního rozdělení (EAP). Pro přístup založený na Bayesovském modálním odhadu (MAP) je volena možnost method = "MAP", pro přístup založený na maximální věrohodnosti je volena možnost method = "ML".*

```
write.csv(theta, "score_MULTIDIM_EAP.csv")
```

*# Příkaz vede k uložení hodnot úrovně zvládnutí hodnocených konstruktů žáky ( $\theta$ ) v csv souboru score\_MULTIDIM\_EAP.*

**Přístup CTT/IRT pro datový rámec TACR\_DATA – alternativní bodová škála**

```
MULTIDIMmodel <- mirt(TACR_DATA, 1, itemtype = "2PL")
```

```
theta <- fscores(MULTIDIMmodel, method = "EAP")
```

*# Tyto dva příkazy replikují odhad multidimenzionálního modelu s tím, že v tomto případě je voleno řešení s jednou dimenzí (konstruktem, faktorem). Výstupem příkazů jsou hodnoty úrovně zvládnutí hodnocených konstruktů žáky ( $\Theta$ ).*

```
theta_transform <- score.transform(theta, 500, 100)
```

*# Funkce score.transform (R balíček CTT) vede k transformaci hodnot úrovně zvládnutí hodnoceného konstruktů žáky ( $\Theta$ ) uložených v objektu theta na alternativní hodovou škálu s průměrem 500 a směrodatnou odchylkou 100. Hodnoty úrovně zvládnutí hodnoceného konstruktů žáky ( $\Theta$ ) na alternativní bodové škále jsou uloženy v objektu theta\_transform. Součástí tohoto objektu je také hodnota percentilového pořadí pro každého žáka.*

### 3.3.10 Propojení výsledků žáků ve dvou ověřovacích testech se společnými (kotvícími) testovými položkami na stejnou škálu

Název modelové situace	
Propojení výsledků žáků ve dvou ověřovacích testech se společnými (kotvícími) testovými položkami na stejnou škálu	
Metodický postup řešení modelové situace	
Metodický postup řešení modelové situace sleduje cíl propojit výsledky žáků ve dvou ověřovacích testech se společnými (kotvícími) položkami na stejnou škálu, přičemž předpokládáme dva neekvivalentní výběrové soubory testovaných žáků (NEAT přístup ke sběru dat). Postup řešení modelové situace se skládá z následujících kroků uživatele metodiky:	
Krok 1: Uživatel metodiky vybírá, zda bude hodnotit výsledky žáků na škále s využitím odhadu modelu vycházejícího z IRT.	
<i>Ne (přístup CTT)</i>	<i>Ano (přístup IRT)</i>
Krok 2: Uživatel metodiky stanoví skóre žáků v řešení testových položek prvního testu a skóre těchto žáků v řešení kotvících testových položek (kotvící test).	Krok 2: Uživatel metodiky odhaduje parametry vybraného modelu prvního testu v souladu s metodickým postupem modelové situace „Volba a využití škály pro stanovení výsledků žáků v ověřovacím testu“.
Krok 3: Uživatel metodiky stanoví skóre žáků v řešení testových položek druhého testu a skóre těchto žáků v řešení kotvících testových položek (kotvící test).	Krok 3: Uživatel metodiky odhaduje parametry vybraného modelu 1 druhého testu v souladu s metodickým postupem modelové situace „Volba a využití škály pro stanovení výsledků žáků v ověřovacím testu“.
Krok 4: Uživatel metodiky počítá četnost výskytu dvojic hodnot dosaženého skóre žáků v řešení prvního testu a dosaženého skóre těchto žáků v řešení kotvícího testu.	

<p>Krok 5: Uživatel metodiky počítá četnost výskytu dvojic hodnot dosaženého skóre žáků v řešení druhého testu a dosaženého skóre těchto žáků v řešení kotvícího testu.</p> <p>Krok 6: Uživatel metodiky vybírá metodu propojení skóre žáků obou testů na stejnou škálu, přičemž může prostřednictvím stanovení váhy zohlednit různou velikost výchozích populací výběrových souborů žáků</p> <hr/> <p><i>Krok 6a): Ekvipercentilní přístup k propojení škál obou testů</i></p> <hr/> <p><i>Krok 6b): Lineární přístup k propojení škál obou testů</i></p>	<p>Krok 4: Uživatel metodiky transformuje škálu jednoho z testů nalezením hodnot parametrů <math>A^*</math> a <math>B^*</math> lineárního vztahu <math>\theta_{ij} = A^* \theta_{il} + B^*</math>, kde index <math>i</math> označuje žáka, index <math>J</math> označuje škálu druhého testu a index <math>I</math> škálu prvního testu.</p> <p>Krok 5: Uživatel metodiky propojuje skóre žáků prvního a druhého testu, a to s využitím informací o úrovni zvládnutí hodnoceného konstruktů žáky (<math>\theta</math>) na stejné škále. Prostřednictvím stanovení váhy může uživatel metodiky zohlednit různou velikost výchozích populací výběrových souborů žáků.</p>
<b>Výstup řešení modelové situace</b>	
Výstupem řešení modelové situace je dvojice srovnatelných skór žáků na společné škále.	Výstupem řešení modelové situace je dvojice srovnatelných skór žáků na společné škále.
<b>Širší situační kontext</b>	
<p>Modelová situace vychází ze sběru dat, který je založen na dvou neekvivalentních výběrových souborech žáků (vzhledem k populaci i vzhledem k úrovni zvládnutí hodnoceného konstruktů) a dvou ověřovacích testech, které obsahují určitý počet stejných, tj. kotvících testových položek. Přístupy vycházející ze sběru dat pro jeden výběrový soubor žáků, respektive pro dva ekvivalentní výběrové soubory žáků jsou jednodušší, neboť umožňují přímé propojení výsledků žáků na společnou škálu dvou ověřovacích testů.</p> <p>S ohledem na důležitost kotvících testových položek pro řešení modelové situace se dvěma neekvivalentními výběrovými soubory žáků je na tyto položky kladena řada požadavků, kdy v ideálním případě by měl být kotvící test mini-verzí obou propojovaných ověřovacích testů (např. široké spektrum obtížností a formát kotvících testových položek) a zároveň by měly kotvící položky „fungovat“ v obou ověřovacích testech podobně (např. umístění kotvících testových položek, vysoká korelace kotvících testových položek a skóre obou ověřovacích testů). Doporučováno je, aby nejméně 20 % testových položek bylo součástí kotvícího testu. Pro propojování výsledků žáků na společnou škálu pak je obecná potřeba dostatečně velkých výběrových souborů žáků, pro přístupy vycházející z IRT je doporučována minimální velikost kolem tří tisíce žáků.</p> <p>Modelová situace je relevantní pro řešení řady praktických úkolů, přičemž vedle propojování výsledků žáků ve dvou ověřovacích testech na jejich společnou škálu je možné metodický postup rozšířit také na propojování vyššího počtu dílčích verzí ověřovacích testů.</p>	
<b>Software a ilustrace jeho využití pro řešení modelové situace</b>	
R balíček <i>equate</i> (viz Albano, 2018)	R balíček <i>equateIRT</i> (viz Battauz, 2018) R balíček <i>ltm</i> (viz Rizopoulos, 2018)



**Přístup CTT pro datové rámce TACR\_DATA\_EQUATE1 a TACR\_DATA\_EQUATE2**

```
FREQ1 <- freqtab(TACR_DATA_EQUATE1, scales = list(0:38, 0:20))
```

*# Funkce freqtab (R balíček equate) vede k výpočtu četností výskytu dvojic skór dosažených žáky v řešení: (a) celého prvního testu; a (b) kotvícího testu, přičemž četnosti dvojic skór jsou uloženy v objektu FREQ1.*

*# Atribut scales uvádí rozsah možných skór žáků v řešení: (a) celého prvního testu (0 až 38); a (b) kotvícího testu (0 až 20).*

```
FREQ2 <- freqtab(TACR_DATA_EQUATE2, scales = list(0:38, 0:20))
```

*# Funkce freqtab (R balíček equate) vede k výpočtu četností výskytu dvojic skór dosažených žáky v řešení: (a) celého druhého testu; a (b) kotvícího testu, přičemž četnosti dvojic skór jsou uloženy v objektu FREQ2.*

*# Atribut scales uvádí rozsah možných skór žáků v řešení: (a) celého prvního testu (0 až 38); a (b) kotvícího testu (0 až 20).*

```
plot(FREQ1); plot(FREQ1)
```

*# Funkce plot (R balíček equate) umožňuje zobrazit četností výskytu dvojic skór žáků v řešení: (a) celého testu; a (b) kotvícího testu graficky.*

```
EQUATE <- equate(FREQ1, FREQ2, type = "equipercentile", method = "frequency estimation", ws = 0.50)
```

*# Funkce equate (R balíček equate) vede k výpočtu korespondujících hodnot skór žáků řešících první a druhý test. Svou roli hraje v tomto ohledu pořadí uvedení objektů FREQ1 a FREQ2, kdy se skóre žáků řešících první test promítají na škálu žáků řešících druhý test. Korespondující hodnoty jsou ukládány v objektu EQUATE.*

*# Atributy type a method vyjadřují, jakým způsobem budou skóre prvního a druhého testu převedena na společnou škálu (zde ekvipercentilní přístup založený na odhadu četností).*

*# Atribut ws zohledňuje váhu populace výběrového souboru žáků řešících první test v syntetické populaci obou výběrových souborů žáků. Zde je váha stanovena shodně pro obě populace výběrových souborů žáků (0.50).*

```
EQUATE$concordance
```

*# Příkaz vede k „vytažení“ a zobrazení dvojic korespondujících skór první a druhé verze testu z objektu EQUATE.*

**Přístup IRT pro datové rámce TACR\_DATA a TACR\_DATA\_ALTER a odhad 3PL modelu**

```
PL3model <- tpm(TACR_DATA)
```

*# Funkce tpm (R balíček ltm) vede k odhadu parametrů 3PL modelu pro datový rámec TACR\_DATA, které jsou uloženy v objektu PL3model.*

```
PL3model_alter <- tpm(TACR_DATA_ALTER)
```

*# Funkce tpm (R balíček ltm) vede k odhadu parametrů 3PL modelu pro datový rámec TACR\_DATA\_ALTER, které jsou uloženy v objektu PL3model\_alter.*

```
PL3model_estm <- import.ltm(PL3model)
```

*# Funkce import.ltm (R balíček equateIRT) vede k „vytažení“ parametrů testových položek a matice jejich kovariancí z objektu PL3model a k jejich uložení v objektu PL3model\_estm.*

```
PL3model_estm_alter <- import.ltm(PL3model_alter)
```

*# Funkce import.ltm (R balíček equateIRT) vede k „vytažení“ parametrů testových položek a matice jejich kovariancí z objektu PL3model\_alter a k jejich uložení v objektu PL3model\_estm\_alter.*

```
PL3model_estc <- list(PL3model_estm$coef, PL3model_estm_alter$coef)
```

```
PL3model_estv <- list(PL3model_estm$var, PL3model_estm_alter$var)
```

*# Funkce list (R balíček equateIRT) vede k vytvoření dvou objektů, které souhrnně obsahují parametry testových položek (PL3model\_estc) a matice kovariancí mezi nimi (PL3model\_estv) základního a alternativního testu. Tyto jsou „vytaženy“ z objektů PL3model\_estm a PL3model\_estm\_alter.*

```
test <- paste("test", 1:2, sep = "")
```

*# Funkce paste vede k vytvoření objektu test pro označení testů jako test1 a test2.*

```
modPL3model <- modIRT(coef = PL3model_estc, var = PL3model_estv, names = test)
```

```
PL3model_I12 <- direc(mods = modPL3model, which = c(1,2), method = "Haebara")
```

*# Funkce modIRT (balíček equateIRT) a direc (balíček equateIRT) vedou k převedení parametrů dvou odhadovaných modelů na společnou škálu, a to s využitím objektů obsahujících hodnoty parametrů testových položek obou modelů (atribut coef = PL3model\_estc), respektive kovariancí mezi nimi (atribut var = PL3model\_estv). Hlavním výstupem využití obou funkcí jsou hodnoty  $A^*$  a  $B^*$  pro transformaci parametrů obou modelů na společnou škálu. Pro jejich zobrazení lze využít obecnou funkci summary pro objekt PL3model\_I12.*

*# Atribut names je využit pro označení obou testů jako test1 a test2.*

*# Atribut which specifikuje, které testy jsou propojovány (zde test1 a test2).*

*# Atribut method specifikuje, který metodický přístup je pro propojování škál testů využit (zde Haeberův přístup).*

```
score(PL3model_I12, method = "OSE", w = 0.50)
```

*# Funkce score (balíček equateIRT) vede k odhadu a zobrazení korespondujících si dvojic skór základního a alternativního testu.*

*# Atribut method specifikuje, který metodický přístup je pro propojování škál testů využit (zde přístup založený na pozorovaných datech).*

*# Atribut w zohledňuje váhu populace výběrového souboru žáků řešících první test v syntetické populaci obou výběrových souborů žáků. Zde je váha stanovena shodně pro obě populace výběrových souborů žáků (0.50).*

## 3.3.11 Pokrok ve vzdělávání s využitím ověřovacích testů

Název modelové situace									
Pokrok ve vzdělávání s využitím ověřovacích testů									
Metodický postup řešení modelové situace									
<p>Záměrem metodického postupu řešení modelové situace je vyhodnotit dosažený pokrok ve vzdělávání v čase, a to s možností hodnocení na různých úrovních vzdělávacího systému (vzdělávací systém, škola, třída/učitel, žák). Takto je předpokladem řešení modelové situace dostupnost dat o vzdělávacích výsledcích stejných žáků ve více časových okamžicích (např. ověřovací testy ve více ročnících studia). Postup řešení modelové situace pro dva ověřovací testy se skládá z následujících kroků uživatele metodiky:</p> <p>Krok 1: Uživatel metodiky vybírá s ohledem na úroveň hodnocení model, kterým bude hodnotit pokrok ve vzdělávání prostřednictvím ověřovacích testů.</p> <table border="1"> <thead> <tr> <th><i>Model přidané hodnoty</i></th> <th><i>Žákovský percentil růst</i></th> </tr> </thead> <tbody> <tr> <td>Krok 2: Uživatel metodiky kompiluje data vztahující se: (a) ke vzdělávacím výsledkům žáků v alespoň dvou ověřovacích testech; (b) k charakteristikám žáků; a (c) k charakteristikám třídy/učitele a školy.</td> <td>Krok 2: Uživatel metodiky vybírá skupinu žáků, kteří dosáhli podobného výsledku jako hodnocený žák v některém z předchozích ověřovacích testů.</td> </tr> <tr> <td>Krok 3: Uživatel metodiky odhaduje hierarchický regresní model vysvětlující výsledek žáků v ověřovacím testu proměnnými na úrovni žáka, třídy/učitele a školy.</td> <td>Krok 3: Uživatel metodiky stanovuje percentil žáka v novém testu, a to v rámci skupiny žáků, která dosáhla podobného výsledku jako hodnocený žák v některém z předchozích ověřovacích testů.</td> </tr> <tr> <td>Krok 4: Uživatel metodiky získává informace o regresních koeficientech relevantních pro jeho záměr hodnocení.</td> <td>Krok 4: Uživatel metodiky agreguje výsledky na vyšší úrovni, než je úroveň žáka (třída/učitel, škola).</td> </tr> </tbody> </table>		<i>Model přidané hodnoty</i>	<i>Žákovský percentil růst</i>	Krok 2: Uživatel metodiky kompiluje data vztahující se: (a) ke vzdělávacím výsledkům žáků v alespoň dvou ověřovacích testech; (b) k charakteristikám žáků; a (c) k charakteristikám třídy/učitele a školy.	Krok 2: Uživatel metodiky vybírá skupinu žáků, kteří dosáhli podobného výsledku jako hodnocený žák v některém z předchozích ověřovacích testů.	Krok 3: Uživatel metodiky odhaduje hierarchický regresní model vysvětlující výsledek žáků v ověřovacím testu proměnnými na úrovni žáka, třídy/učitele a školy.	Krok 3: Uživatel metodiky stanovuje percentil žáka v novém testu, a to v rámci skupiny žáků, která dosáhla podobného výsledku jako hodnocený žák v některém z předchozích ověřovacích testů.	Krok 4: Uživatel metodiky získává informace o regresních koeficientech relevantních pro jeho záměr hodnocení.	Krok 4: Uživatel metodiky agreguje výsledky na vyšší úrovni, než je úroveň žáka (třída/učitel, škola).
<i>Model přidané hodnoty</i>	<i>Žákovský percentil růst</i>								
Krok 2: Uživatel metodiky kompiluje data vztahující se: (a) ke vzdělávacím výsledkům žáků v alespoň dvou ověřovacích testech; (b) k charakteristikám žáků; a (c) k charakteristikám třídy/učitele a školy.	Krok 2: Uživatel metodiky vybírá skupinu žáků, kteří dosáhli podobného výsledku jako hodnocený žák v některém z předchozích ověřovacích testů.								
Krok 3: Uživatel metodiky odhaduje hierarchický regresní model vysvětlující výsledek žáků v ověřovacím testu proměnnými na úrovni žáka, třídy/učitele a školy.	Krok 3: Uživatel metodiky stanovuje percentil žáka v novém testu, a to v rámci skupiny žáků, která dosáhla podobného výsledku jako hodnocený žák v některém z předchozích ověřovacích testů.								
Krok 4: Uživatel metodiky získává informace o regresních koeficientech relevantních pro jeho záměr hodnocení.	Krok 4: Uživatel metodiky agreguje výsledky na vyšší úrovni, než je úroveň žáka (třída/učitel, škola).								
Výstup řešení modelové situace									
Výstupem řešení modelové situace jsou informace o regresních koeficientech proměnných vstupujících do hierarchického regresního modelu jako proměnné vysvětlující. Tyto mohou zahrnovat standardizované hodnoty regresního koeficientu, stejně jako informaci o jejich statistické významnosti.	Základním výstupem řešení modelové situace je percentil žáka v novém ověřovacím testu. Vysoké hodnoty percentilu naznačují vyšší pokrok žáka ve vzdělávání, naopak nízké hodnoty naznačují omezený pokrok žáka ve vzdělávání. Možná je další agregace hodnot na vyšší úroveň hodnocení.								
Širší situační kontext									
<p>Volba modelu hodnocení pokroku ve vzdělávání je ovlivněna řadou souvisejících aspektů, jako je záměr hodnocení či povaha dat, například:</p> <ul style="list-style-type: none"> <li>Pro hodnocení vzdělávacího pokroku žáka je vhodný metodický postup sledující model žakovského percentilu růstu a nikoliv metodický postup sledující model přidané hodnoty.</li> </ul>									

- Pro hodnocení vzdělávacího pokroku žáků na úrovni systému je vhodný metodický postup sledující model přidané hodnoty a nikoliv metodický postup sledující model žakovského percentilu růstu. Hodnocení je v tomto případě založeno na informacích o regresním koeficientu, který se vztahuje k výsledkům žáků v předchozích ověřovacích testováních.

Metodický postup modelu přidané hodnoty je založený na odhadech hierarchických regresních modelů, které jsou navíc dále komplikovány přítomností autokorelace utvářené korelací mezi „sousedícími“ daty v časové řadě. Z tohoto důvodu není komplexní problematika odhadů takových modelů v této metodice přímo řešena.

### 3.3.12 Unidimenzionalita ověřovacího testu a počet konstruktů v něm obsažených

Název modelové situace
Unidimenzionalita ověřovacího testu a počet konstruktů v něm obsažených
Metodický postup řešení modelové situace
<p>Záměrem řešení modelové situace je vyhodnotit naplnění předpokladu unidimenzionality (výskytu jen jednoho hlavního hodnoceného konstruktů) ověřovacího testu a v případě narušení tohoto předpokladu stanovit optimální počet konstruktů v testu obsažených a současně také sílu konstruktů hlavního. Významnou motivací uživatele metodiky ke sledování tohoto postupu je ta skutečnost, že silné narušení předpokladu unidimenzionality ověřovacího testu zpochybňuje kvalitu dalších odhadů, které jsou na předpokladu unidimenzionality ověřovacího testu založeny (např. odhad tradičních modelů vycházejících z IRT).</p> <p>Pro řešení modelové situace může uživatel metodiky využít řadu různých metodických přístupů, které jsou zachyceny v následujících krocích:</p> <p>Krok 1: Uživatel metodiky vybírá metodický přístup pro hodnocení unidimenzionality ověřovacího testu.</p> <hr/> <p><i>Krok 1a): Uživatel metodiky počítá s ohledem na dichotomický charakter testových položek hodnoty tetrachorických korelací mezi nimi.</i></p> <hr/> <p>Krok 2a): Uživatel metodiky posuzuje vztahy v matici tetrachorických korelací, přičemž se zaměřuje na identifikaci a interpretaci vztahů testových položek s vysokými hodnotami tetrachorických korelací.</p> <hr/> <p><i>Krok 1b): Uživatel metodiky počítá hodnoty vlastních čísel (eigenvalues) faktorů (konstruktů, dimenzí), přičemž za tímto účelem vybírá vhodnou metodu faktorové analýzy. Doporučena je preference metody hlavních os.</i></p> <hr/> <p>Krok 2b): Uživatel metodiky volí optimální počet faktorů (konstruktů, dimenzí) prostřednictvím Kaiserova kritéria, kdy doporučený počet faktorů (konstruktů, dimenzí) odpovídá počtu faktorů s vlastním číslem vyšším než jedna.</p> <p>Krok 3b): Uživatel metodiky volí optimální počet faktorů (konstruktů, dimenzí) na základě významného zlomu v sutinovém grafu (<i>scree plot</i>), kdy doporučený počet faktorů (konstruktů, dimenzí) je umístěn nad významným zlomem v grafu.</p>

Krok 4b): Uživatel metodiky volí optimální počet faktorů (konstruktů, dimenzí) s využitím metody paralelní analýzy, která srovnává skutečné hodnoty vlastních čísel a simulované hodnoty vlastních čísel. Doporučený počet faktorů (konstruktů, dimenzí) je dán počtem faktorů, jejichž skutečná hodnota vlastního čísla je vyšší než simulovaná hodnota vlastního čísla.

*Krok 1c): Uživatel metodiky počítá hodnoty VSS kritéria (metoda velmi jednoduché struktury) pro jím stanovený maximální počet faktorů a s rozlišením VSS komplexity 1 a více komplexní VSS komplexity 2.*

Krok 2c): Uživatel metodiky volí optimální počet faktorů (konstruktů, dimenzí) podle nejvyšší hodnoty VSS kritéria komplexity 1 či 2.

*Krok 1d): Uživatel metodiky počítá průměrné hodnoty korelací mimo hlavní diagonálu (MAP), přičemž v každém kroku odstraňuje vliv nejvýznamnějšího konstruktů (komponenty). Postup je opakován v počtu kroků, které odpovídají počtu testových položek (tzv. Velicerův MAP test).*

Krok 2d): Uživatel metodiky volí doporučený počet faktorů (konstruktů, dimenzí) podle nejnižší hodnoty MAP.

*Krok 1e): Uživatel metodiky hodnotí sílu narušení předpokladu unidimenzionality testu, tj. úroveň multidimenzionality testu, prostřednictvím výpočtu DETECT indexu.*

Krok 2e): Uživatel metodiky posuzuje sílu narušení předpokladu unidimenzionality testu na následující škále:

- Hodnoty nižší než 0,1 (0,2) naznačují unidimenzionalitu testu.
- Hodnoty v intervalu 0,1 až 0,5 (0,2 až 0,4) naznačují slabou multidimenzionalitu testu.
- Hodnoty v intervalu 0,5 až 1,0 (0,4 až 1,0) naznačují středně silnou multidimenzionalitu testu.
- Hodnoty vyšší než 1,0 naznačují silnou multidimenzionalitu testu.

*Krok 1f): Uživatel metodiky počítá ukazatele spolehlivosti testu  $\omega_h$  a  $\omega_b$ , kde ukazatel  $\omega_h$  se vztahuje jen k hlavnímu faktoru (konstruktů, dimenzí) a ukazatel  $\omega_b$  zohledňuje také vliv skupinových faktorů (konstruktů, dimenzí) dílčích testových položek.*

Krok 2f): Pro další interpretaci výsledků viz modelová situace „Spolehlivost (škály) ověřovacího testu“

#### Výstup řešení modelové situace

Výstupem metodického přístupu a) jsou hodnoty tetrachorických korelací pro všechny dvojice testových položek. Shluky vysokých hodnot utváří podezření na narušení předpokladu unidimenzionality ověřovacího testu.

Výstupem metodického přístupu b) jsou hodnoty optimálního počtu konstruktů (faktorů, dimenzí) jednotlivých metodických přístupů (Kaiserovo kritérium, sutinový graf, paralelní analýza). Naplnění předpokladu unidimenzionality ověřovacího testu je zde spojeno s optimálním počtem jednoho konstruktů (faktoru, dimenze).

Výstupem metodického přístupu c) jsou hodnoty optimálního počtu konstruktů (faktorů, dimenzí) jednotlivých metodických přístupů (VSS komplexita 1 a VSS komplexita 2). Naplnění předpokladu unidimenzionality ověřovacího testu je zde spojeno s optimálním počtem jednoho konstruktů (faktoru, dimenze).

Výstupem metodického přístupu d) je hodnota optimálního počtu konstruktů (faktorů, dimenzí) podle hodnot MAP (Velicerův MAP test). Naplnění předpokladu unidimenzionality ověřovacího testu je zde spojeno s optimálním počtem jednoho konstruktů (faktoru, dimenze).

Výstupem metodického přístupu e) je hodnota DETECT indexu. Pro naplnění předpokladu unidimenzionality ověřovacího testu jsou žádoucí hodnoty DETECT indexu, které nenaznačují silnou multidimenzionalitu testu.

Výstupem metodického přístupu f) jsou hodnoty ukazatelů spolehlivosti ověřovacího testu  $\omega_h$  a  $\omega_t$ . Vysoké hodnoty ukazatele  $\omega_t$  ukazují na možný vliv skupinových konstruktů (faktorů, dimenzí) dílčích testových položek narušujících předpoklad unidimenzionality ověřovacího testu.

#### Širší situační kontext

Hodnocení naplnění předpokladu unidimenzionality ověřovacího testu poskytuje přínosné informace pro poznání konstruktů, které ověřovací test měří, a to také konstrukty dané: (a) stejným uvozujícím textem testových položek; (b) nespravedlivým chováním testových položek vůči určité skupině žáků (DIF analýza); či (c) umístěním testových položek na konci ověřovacího testu. Existence takto utvářených konstruktů může být ukázána také hodnocením lokální nezávislosti testových položek

Hodnocení naplnění předpokladu unidimenzionality ověřovacího testu může poskytnout novou informaci o konstruktech v testu obsažených, ale také potvrdit ex-ante očekávání, která vzešla například z expertního posouzení. Alternativně lze aplikovat přístup založený na klastrové analýze testových položek ověřovacího testu.

Hodnocení unidimenzionality ověřovacího testu je významným nástrojem pro rozhodnutí, jakou škálu pro měření úrovně zvládnutí hodnoceného konstruktů zvolit, například zda preferovat přístupy, které vycházejí z CTT před přístupy, které vycházejí z IRT, případně zda dát přednost odhadu multidimenzionálních modelů.

Při narušení předpokladu unidimenzionality ověřovacího testu je potřeba s opatrností interpretovat řadu odhadů, především pak odhad hodnot ukazatelů spolehlivosti (např. Cronbachovo alfa).

V praxi se zjevně s dokonalými unidimenzionálními testy nesečkáme, což je žádoucí vzít do úvahy při rozhodování o počtu konstruktů v ověřovacím testu obsažených. Vhodným nástrojem, který zohledňuje tuto myšlenku, je DETECT index.

Z praktického hlediska se jako vhodné jeví diskutovat závěry různých metodických přístupů k hodnocení unidimenzionality ověřovacího testu a ke stanovení optimálního počtu konstruktů v něm obsažených. Takto je například paralelní analýza citlivá k nadhodnocení optimálního počtu faktorů pro velké výběrové soubory žáků.

#### Software a ilustrace jeho využití pro řešení modelové situace

R balíček *psych* (viz Revelle, 2020); R balíček *sirt* (viz Robitzsch, 2020)

**Tetrachorické korelace pro datový rámec TACR\_DATA**

```
tetrakorelace <- mixed.cor(TACR_DATA)$rho
```

*# Funkce mixed.cor (R balíček psych) vede k výpočtu korelací, jejichž typ je stanoven na základě povahy vstupních dat, přičemž korelace jsou „vytaženy“ a uloženy v objektu tetrakorelace.*

```
cor.plot(tetrakorelace)
```

*# Funkce cor.plot (R balíček psych) vede k vykreslení grafu korelací dvojic testových položek uložených v objektu tetrakorelace.*

**Metodické postupy založené na výpočtu vlastních čísel pro datový rámec TACR\_DATA**

```
TACR_DATA_paralel <- fa.parallel(TACR_DATA, fm="ml", fa="fa", cor="tet")
```

*# Funkce fa.parallel (R balíček psych) vede jednak k vykreslení sutinového grafu (scree plot), jednak k výpočtu vlastních čísel faktorů faktorové analýzy datového rámce TACR\_DATA a jednak k zobrazení optimálního počtu konstruktů (faktorů, dimenzí) paralelní analýzou. Výsledky jsou uloženy v objektu TACR\_DATA\_paralel.*

*# Atributy fm a fa umožňují výběr podoby faktorové analýzy, zde je volena metoda maximální věrohodnosti (fm = "ml") a metoda hlavních osob (fa = "fa").*

*# Atribut cor umožňuje vybrat typ korelací, které jsou využity ve faktorové analýze, zde tetrachorické korelace.*

```
TACR_DATA_paralel$fa.values
```

*# Příkaz vede k „vytažení“ a zobrazení hodnot vlastních čísel (eigenvalues) z objektu TACR\_DATA\_paralel.*

Tímto způsobem získáváme všechny potřebné informace pro řešení modelové situace prostřednictvím Kaiserova kritéria, sutinového grafu a metody paralelní analýzy.

**VSS kritéria a Velicerův MAP test pro datový rámec TACR\_DATA**

```
vss(TACR_DATA, n=38, fm="ml", cor="tet")
```

*# Funkce vss (R balíček psych) vede k odhadu hodnot kritérií VSS (VSS komplexity 1 a VSS komplexity 2) a MAP pro počet faktorů 1 až 38 (atribut n = 38). Zároveň je pro metody velmi jednoduché struktury (VSS) a Velicerova MAP testu stanoven optimální počet konstruktů (faktorů, dimenzí).*

*# Atribut fm umožňuje výběr podoby faktorové analýzy, zde je volena metoda maximální věrohodnosti.*

*# Atribut cor umožňuje vybrat typ korelací, které jsou využity ve faktorové analýze, zde tetrachorické korelace.*

Tímto způsobem získáváme všechny potřebné informace pro řešení modelové situace prostřednictvím metody velmi jednoduché struktury (VSS) a Velicerova MAP testu.

**DETECT index pro datové rámce TACR\_DATA a DATA\_TACR\_HIERCH**

```
Skore <- TACR_DATA_HIEARCH$skore
```

*# Příkaz vede k „vytažení“ hodnot skóre z datového rámce TACR\_DATA\_HIEARCH a uložení v objektu Skore.*

```
expl.detect(TACR_DATA, Skore, nclusters=38)
```

# Funkce `expl.detect` (R balíček `sirt`) vede k výpočtu a zobrazení hodnoty `DETECT` indexu pro počet faktorů/klastrů 1 až 38 (atribut `nclusters = 38`).

### 3.3.13 Lokální nezávislost testových položek

Název modelové situace					
Lokální nezávislost testových položek					
Metodický postup řešení modelové situace					
<p>Záměrem řešení modelové situace je ověřit naplnění předpokladu lokální nezávislosti testových položek ověřovacího testu. Motivací uživatele metodiky je v tomto ohledu zejména ta skutečnost, že narušení předpokladu lokální nezávislosti testových položek narušuje kvalitu dalších odhadů, a to především tradičních modelů vycházejících z IRT. Postup řešení modelové situace pro posouzení lokální nezávislosti testových položek ověřovacího testu se skládá z následujících kroků uživatele metodiky:</p> <p>Krok 1: Uživatel metodiky vybírá, zda bude hodnotit výsledky žáků na škále s využitím modelu vycházejícího z IRT.</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: center;"><i>Ne (přístup CTT)</i></th> <th style="text-align: center;"><i>Ano (přístup IRT)</i></th> </tr> </thead> <tbody> <tr> <td style="vertical-align: top;"> <p>Krok 2: Uživatel metodiky počítá s ohledem na dichotomický charakter testových položek hodnoty tetrachorických korelací mezi nimi.</p> </td> <td style="vertical-align: top;"> <p>Krok 2: Uživatel metodiky odhaduje 1PL model v souladu s metodickým postupem modelové situace „Volba a využití škály pro stanovení výsledků žáků v ověřovacím testu“.</p> <p>Krok 3: Uživatel metodiky počítá s využitím odhadu parametrů modelu statistiku Q3, a to pro každou dvojici testových položek.</p> </td> </tr> </tbody> </table>		<i>Ne (přístup CTT)</i>	<i>Ano (přístup IRT)</i>	<p>Krok 2: Uživatel metodiky počítá s ohledem na dichotomický charakter testových položek hodnoty tetrachorických korelací mezi nimi.</p>	<p>Krok 2: Uživatel metodiky odhaduje 1PL model v souladu s metodickým postupem modelové situace „Volba a využití škály pro stanovení výsledků žáků v ověřovacím testu“.</p> <p>Krok 3: Uživatel metodiky počítá s využitím odhadu parametrů modelu statistiku Q3, a to pro každou dvojici testových položek.</p>
<i>Ne (přístup CTT)</i>	<i>Ano (přístup IRT)</i>				
<p>Krok 2: Uživatel metodiky počítá s ohledem na dichotomický charakter testových položek hodnoty tetrachorických korelací mezi nimi.</p>	<p>Krok 2: Uživatel metodiky odhaduje 1PL model v souladu s metodickým postupem modelové situace „Volba a využití škály pro stanovení výsledků žáků v ověřovacím testu“.</p> <p>Krok 3: Uživatel metodiky počítá s využitím odhadu parametrů modelu statistiku Q3, a to pro každou dvojici testových položek.</p>				
Výstup řešení modelové situace					
<p>Výstupem řešení modelové situace jsou hodnoty tetrachorických korelací pro všechny dvojice testových položek ověřovacího testu. Vysoké hodnoty tetrachorických korelací naznačují možné narušení lokální nezávislosti testových položek.</p>	<p>Výstupem řešení modelové situace jsou hodnoty statistiky Q3 pro všechny dvojice testových položek ověřovacího testu. Vysoké hodnoty statistiky Q3 naznačují možné narušení lokální nezávislosti testových položek.</p>				
Širší situační kontext					
<p>Hodnocení lokální nezávislosti testových položek může mít vztah k hodnocení unidimenzionality ověřovacího testu, neboť silně korelované testové položky potenciálně utváří vedlejší konstrukt ověřovacího testu, který může být také dán: (a) stejným uvozujícím textem testových položek; (b) nespravedlivým chováním testových položek vůči určité skupině žáků (DIF); či (c) umístěním testových položek na konci ověřovacího testu.</p>					



Při narušení předpokladu lokální nezávislosti testových položek ověřovacího testu je možné zvážit opatření, které povedou k řešení tohoto problému (např. převedení dichotomických testových položek na polytomické). V opačném případě může mít narušení předpokladu lokální nezávislosti testových položek nepříznivý vliv na řešení modelových situací, které lokální nezávislost testových položek předpokládají (např. odhad tradičních modelů vycházejících z IRT).

### Software a ilustrace jeho využití pro řešení modelové situace

R balíček *psych* (viz Revelle, 2020)

R balíček *sirt* (viz Robitzsch, 2020)

#### Přístup CTT pro datový rámec TACR\_DATA

```
tetrakorelace <- mixed.cor(TACR_DATA)$rho
```

*# Funkce mixed.cor (R balíček psych) vede k výpočtu korelací, jejichž typ je stanoven na základě povahy vstupních dat, přičemž korelace jsou „vytaženy“ a uloženy v objektu tetrakorelace.*

```
cor.plot(tetrakorelace)
```

*# Funkce cor.plot (R balíček psych) vede k vykreslení grafu korelací dvojic testových položek uložených v objektu tetrakorelace.*

#### Přístup IRT pro datový rámec TACR\_DATA

```
PL1model <- rasch.mml2(TACR_DATA)
```

*# Funkce rasch.mml2 (R balíček sirt) vede k odhadu parametrů 1PL modelu datového rámce TACR\_DATA a k jejich uložení v objektu PL1model.*

```
PL1model_theta <- wle.rasch(dat = TACR_DATA, b = PL1model$item$b)
```

*# Funkce wle.rasch (R balíček sirt) vede k odhadu hodnot úrovně zvládnutí hodnoceného konstruktů  $\theta$  pro odhad PL1 modelu datového rámce TACR\_DATA a k jejich uložení v objektu PL1model\_theta.*

*# Atribut b „vytahuje“ hodnoty obtížnosti testových položek b z objektu PL1model.*

```
PL1model_Q3 <- Q3(dat = TACR_DATA, theta = PL1model_theta, b = PL1model$item$b)
```

*# Funkce Q3 (R balíček sirt) vede k odhadům hodnot statistiky Q3 pro každou dvojici testových položek datového rámce TACR\_DATA, přičemž tyto hodnoty jsou uloženy v objektu PL1model\_Q3.*

*# Atribut theta „vytahuje“ hodnoty úrovně zvládnutí hodnoceného konstruktů  $\theta$  z objektu PL1model\_theta.*

*# Atribut b „vytahuje“ hodnoty obtížnosti testových položek b z objektu PL1model.*

```
P1model_Q3$Q3.matrix
```

*# Příkaz vede k „vytažení“ a zobrazení hodnot statistik Q3 z objektu P1model\_Q3.*

## 3.3.14 Dílčí konstrukty ověřovacího testu

Název modelové situace											
Dílčí konstrukty ověřovacího testu											
Metodický postup řešení modelové situace											
<p>Ověřovací test může vedle hlavního konstruktů v sobě obsahovat také dílčí konstrukty utvářené skupinami proměnných, přičemž tyto dílčí konstrukty mohou být definovány ex-ante expertním posouzením nebo exploračně ex-post. Právě na hodnocení dílčích konstruktů ověřovacího testu se zaměřuje modelová situace. Postup jejího řešení se skládá z následujících kroků uživatele metodiky:</p> <p>Krok 1: Uživatel metodiky volí, zda bude pracovat s předem stanoveným počtem konstruktů (faktorů, dimenzí) ověřovacího testu.</p> <table border="1"> <thead> <tr> <th>Ano</th> <th>Ne</th> </tr> </thead> <tbody> <tr> <td>Krok 2: Uživatel metodiky stanovuje počet faktorů prostřednictvím metodických postupů b), c), d), e) řešení modelové situace „Unidimenzionalita ověřovacího testu a počet konstruktů v něm obsažených“.</td> <td>Krok 2: Uživatel metodiky volí parametry klastrové analýzy pro shlukování testových položek ověřovacího testu</td> </tr> <tr> <td>Krok 3: Uživatel metodiky volí parametry klastrové analýzy pro shlukování testových položek ověřovacího testu.</td> <td>Krok 3: Uživatel metodiky shlukuje testové položky ověřovacího testu s využitím klastrové analýzy a průběh shlukování graficky zobrazuje.</td> </tr> <tr> <td>Krok 4: Uživatel metodiky shlukuje testové položky ověřovacího testu do stanoveného počtu faktorů s využitím klastrové analýzy a průběh shlukování graficky zobrazuje.</td> <td>Krok 4: Uživatel metodiky analyzuje grafické znázornění průběhu shlukování a hledá nejlepší, interpretovatelné řešení.</td> </tr> <tr> <td>Krok 5: Uživatel metodiky interpretuje testové položky ověřovacího testu zařazené do stejného shluku.</td> <td>Krok 5: Uživatel metodiky interpretuje zvolené řešení kroku 4 metodického postupu.</td> </tr> </tbody> </table>		Ano	Ne	Krok 2: Uživatel metodiky stanovuje počet faktorů prostřednictvím metodických postupů b), c), d), e) řešení modelové situace „Unidimenzionalita ověřovacího testu a počet konstruktů v něm obsažených“.	Krok 2: Uživatel metodiky volí parametry klastrové analýzy pro shlukování testových položek ověřovacího testu	Krok 3: Uživatel metodiky volí parametry klastrové analýzy pro shlukování testových položek ověřovacího testu.	Krok 3: Uživatel metodiky shlukuje testové položky ověřovacího testu s využitím klastrové analýzy a průběh shlukování graficky zobrazuje.	Krok 4: Uživatel metodiky shlukuje testové položky ověřovacího testu do stanoveného počtu faktorů s využitím klastrové analýzy a průběh shlukování graficky zobrazuje.	Krok 4: Uživatel metodiky analyzuje grafické znázornění průběhu shlukování a hledá nejlepší, interpretovatelné řešení.	Krok 5: Uživatel metodiky interpretuje testové položky ověřovacího testu zařazené do stejného shluku.	Krok 5: Uživatel metodiky interpretuje zvolené řešení kroku 4 metodického postupu.
Ano	Ne										
Krok 2: Uživatel metodiky stanovuje počet faktorů prostřednictvím metodických postupů b), c), d), e) řešení modelové situace „Unidimenzionalita ověřovacího testu a počet konstruktů v něm obsažených“.	Krok 2: Uživatel metodiky volí parametry klastrové analýzy pro shlukování testových položek ověřovacího testu										
Krok 3: Uživatel metodiky volí parametry klastrové analýzy pro shlukování testových položek ověřovacího testu.	Krok 3: Uživatel metodiky shlukuje testové položky ověřovacího testu s využitím klastrové analýzy a průběh shlukování graficky zobrazuje.										
Krok 4: Uživatel metodiky shlukuje testové položky ověřovacího testu do stanoveného počtu faktorů s využitím klastrové analýzy a průběh shlukování graficky zobrazuje.	Krok 4: Uživatel metodiky analyzuje grafické znázornění průběhu shlukování a hledá nejlepší, interpretovatelné řešení.										
Krok 5: Uživatel metodiky interpretuje testové položky ověřovacího testu zařazené do stejného shluku.	Krok 5: Uživatel metodiky interpretuje zvolené řešení kroku 4 metodického postupu.										
Výstup řešení modelové situace											
Výstupem řešení modelové situace jsou shluky souvisejících testových položek ověřovacího testu. Testové položky stejného shluku utváří konstrukt, který uživatel metodiky vhodně interpretuje.	Výstupem řešení modelové situace jsou shluky souvisejících testových položek ověřovacího testu. Testové položky stejného shluku utváří konstrukt, který uživatel metodiky vhodně interpretuje.										
Širší situační kontext											
<p>Klastrová analýza testových položek ověřovacího testu může poskytnout novou informaci o konstruktech v testu obsažených, ale také potvrdit ex-ante očekávání, která vzešla například z expertního posouzení. Alternativně lze vazby testových položek a konstruktů ověřovacího testu hodnotit faktorovou analýzou.</p>											

Software a ilustrace jeho využití pro řešení modelové situace
R balíček <i>psych</i> (viz Revelle, 2020)
<b>Postup řešení pro datový rámec TACR_DATA</b>
tetrakorelace <- mixed.cor(TACR_DATA)\$rho
<i># Funkce mixed.cor (R balíček psych) vede k výpočtu korelací, jejichž typ je stanoven na základě povahy vstupních dat, přičemž korelace jsou „vytaženy“ a uloženy v objektu tetrakorelace.</i>
iclust(tetrakorelace)
<i># Funkce iclust (R balíček psych) vede mimo jiné ke grafickému zobrazení procesu shlukování testových položek datového ráme TACR_DATA.</i>

### 3.3.15 Volba vhodného modelu vycházejícího z IRT pro vyhodnocení ověřovacího testu

Název modelové situace
Volba vhodného modelu vycházejícího z IRT pro vyhodnocení ověřovacího testu
Metodický postup řešení modelové situace
Modelová situace vychází ze záměru uživatele metodiky vyhodnotit ověřovací test s využitím vhodného modelu založeného na IRT. Za tímto účelem je posuzována úroveň shody dat ověřovacího testu s daty, která jsou generována zvažovaným modelem, přičemž žádoucí je dobrá shoda empirických a modelových dat. Rozšířením modelové situace je volba nejvhodnějšího modelu vycházejícího z IRT pro vyhodnocení ověřovacího testu. Preferován je model vykazující nejvyšší úroveň dobré shody empirických a modelových dat. Pokud mezi dvěma modely nejsou v tomto ohledu zaznamenány významné rozdíly, je dána přednost jednoduššímu z nich. Postup řešení modelové situace se následně skládá z následujících kroků uživatele metodiky pro dva výše uvedené záměry:
Záměr 1: Posouzení vhodnosti modelu vycházejícího z IRT pro vyhodnocení ověřovacího testu
Krok 1: Uživatel metodiky odhaduje model vycházející z IRT v souladu s metodickým postupem modelové situace „Volba a využití škály pro stanovení výsledků žáků v ověřovacím testu“.
Krok 2: Uživatel metodiky počítá úroveň dobré shody skutečných a modelem generovaných odpovědí žáků na testové položky ověřovacího testu prostřednictvím: (a) indexu RMSEA; a (b) indexu SRMSR.
Krok 3: Uživatel metodiky srovnává vypočtené hodnoty: (a) indexu RMSEA a (b) indexu SRMSR s kritickou hodnotou 0,05. Vyšší hodnoty zamítají nulovou hypotézu o dobré shodě empirických a modelových dat, a nepotvrzují tak vhodnost posuzovaného modelu pro vyhodnocení ověřovacího testu.
Záměr 2: Volba nejvhodnějšího modelu vycházejícího z IRT pro vyhodnocení ověřovacího testu
Krok 1: Uživatel metodiky odhaduje zvažované modely vycházející z IRT v souladu s metodickým postupem modelové situace „Volba a využití škály pro stanovení výsledků žáků v ověřovacím testu“.

Krok 2: Uživatel metodiky počítá hodnoty indexů dobré shody pro modely odhadované v kroku 1, a to hodnoty:

- (a) poměru věrohodností zahrnutých modelů (LRT);
- (b) Akaikeho informačního kritéria (AIC);
- (c) Bayesova informačního kritéria (BIC).

Krok 3: Uživatel metodiky vybírá jako nejvhodnější model: (a) se statisticky významnou změnou LRT; (b) s nejnižší hodnotou AIC; a (c) s nejnižší hodnotou BIC.

### Výstup řešení modelové situace

Výstupem prvního záměru jsou hodnoty indexů RMSEA a SRMSR, které umožňují posoudit vhodnost posuzovaného modelu vycházejícího z IRT pro vyhodnocení ověřovacího testu.

Výstupem druhého záměru jsou hodnoty tří indexů dobré shody: (a) LRT; (b) AIC; a (c) BIC, které umožňují vybrat nejvhodnější model vycházející z IRT pro vyhodnocení ověřovacího testu.

### Širší situační kontext

Modelová situace úzce navazuje na odhad modelů vycházejících z IRT, neboť pomáhá při výběru nejvhodnějšího z nich. Klíčové principy jsou v tomto ohledu spojeny se zájmem o co nejlepší shodu vybraného modelu s odpověďmi žáků na testové položky ověřovacího testu, ale také o co nejjednodušší model vycházející z IRT.

### Software a ilustrace jeho využití pro řešení modelové situace

R balíček *mirt* (viz Chalmers, 2020); R balíček *ltm* (viz Rizopoulos, 2018)

#### Postup řešení záměru 1 pro datový rámec TACR\_DATA

```
PL2model <- mirt(TACR_DATA, 1)
```

*# Funkce mirt (R balíček mirt) vede k odhadu parametrů 2PL modelu, které jsou uloženy v objektu PL2model.*

*# Atribut 1 vyjadřuje, že má být odhadován unidimenzionální 2PL model.*

```
M2(MirtPL2model)
```

*# Funkce M2 (R balíček mirt) vede k odhadu a zobrazení hodnot indexů RMSEA a SRMSR.*

#### Postup řešení záměru 2 pro datový rámec TACR\_DATA

```
PL1model <- rasch(TACR_DATA)
```

*# Funkce rasch (R balíček ltm) vede k odhadu parametrů 1PL modelu, které jsou uloženy v objektu PL1model.*

```
PL2model <- ltm(TACR_DATA ~ z1)
```

*# Funkce ltm (R balíček ltm) vede k odhadu parametrů 2PL modelu, které jsou uloženy v objektu PL2model.*

*# Parametr z1 odpovídá úrovni zvládnutí hodnoceného konstruktů.*

```

PL3model <- tpm(TACR_DATA)

# Funkce tpm (R balíček ltm) vede k odhadu parametrů 3PL modelu, které jsou uloženy v objektu PL3model.

anova(PL1model, PL2model)

# Funkce anova vede k zobrazení hodnot LRT, AIC a BIC a rovněž statistické významnosti LRT pro objekty PL1model
a PL2model.

anova(PL2model, PL3model)

# Funkce anova vede k zobrazení hodnot LRT, AIC a BIC a rovněž statistické významnosti LRT pro objekty PL2model
a PL3model.

```

### 3.3.16 Základní vyhodnocení a reporting výsledků ověřovacího testu

Název modelové situace
Základní vyhodnocení a reporting výsledků ověřovacího testu
Metodický postup řešení modelové situace
<p>Záměr modelové situace je spojený se základním vyhodnocením a reportingem výsledků ověřovacího testu, tj. vytvořením nejčastějšího výstupu z ověřovacích testů. Za tímto účelem jsou využívány běžné metodické přístupy, přičemž uživatelé metodiky se zároveň nabízejí řada možností, jak výsledky ověřovacího testu vyhodnocovat a reportovat vůči zvoleným proměnným. Postup řešení modelové situace pro vyhodnocení a reporting výsledků ověřovacího testu se skládá z následujících kroků uživatele metodiky:</p> <p>Krok 1: Uživatel metodiky vybírá škálu pro vyhodnocení a reporting výsledků ověřovacího testu.</p> <p>Krok 2: Uživatel metodiky vybírá dílčí tematické části, vůči kterým bude ověřovací test vyhodnocovat a výsledky reportovat.</p> <p>Krok 3: Uživatel metodiky vybírá úroveň vyhodnocení a reportingu výsledků ověřovacího testu s možnostmi: (a) žák; (b) třída/učitel; (c) škola; (d) území; a (e) vzdělávací systém.</p> <p>Krok 4: Uživatel metodiky vybírá dílčí charakteristiky: (a) žáka; (b) třídy; (c) školy; (d) území; a (e) vzdělávacího systému, vůči kterým bude ověřovací test vyhodnocovat a výsledky reportovat.</p> <p>Krok 5: Uživatel metodiky kompiluje datovou matici v souladu se specifikací v krocích 1 až 4 postupu.</p> <p>Krok 6: Uživatel metodiky vyhodnocuje datovou matici kompilovanou v kroku 5 postupu, a to s využitím zvolených metod hodnocení (např. střední hodnota, rozptyl hodnot, četnosti hodnot).</p> <p>Krok 7: Uživatel metodiky reportuje výsledky zjištěné v kroku 6 postupu.</p>
Výstup řešení modelové situace
Výstupem řešení modelové situace je vyhodnocení a reporting hlavních výsledků vyhodnocení ověřovacího testu ve zvolené podobě (např. tabelární podoba, grafická podoba, textová podoba).

Širší situační kontext
<p>Modelová situace má silnou vazbu především k volbě a odhadu nejvhodnější škály pro vyhodnocení a reporting výsledků ověřovacího testu.</p> <p>V případě vyhodnocení a reportingu výsledků ve vazbě na tematické části ověřovacího testu mohou být tyto části utvářeny, či jejich relevance ověřována, s využitím metod faktorové analýzy – identifikace konstruktů (dimenzí, faktorů) ověřovacího testu.</p> <p>K modelové situace má ovšem svůj vztah celá řada dalších modelových situací a metodických postupů (např. propojování škál dvou testů).</p>
Software a ilustrace jeho využití pro řešení modelové situace
<p>Pro řešení modelové situace lze využít běžný software pro práci s datovými soubory (např. MS Excel).</p>

### 3.3.17 Rozdíly ve výsledcích žáků dané rozdíly uvnitř škol a rozdíly mezi školami

Název modelové situace
Rozdíly ve výsledcích žáků dané rozdíly uvnitř škol a rozdíly mezi školami
Metodický postup řešení modelové situace
<p>Záměrem modelové situace je zjistit, do jaké míry jsou rozdíly ve výsledcích žáků v ověřovacím testu utvářeny na úrovni žáka, tj. uvnitř škol, a do jaké míry jsou rozdíly ve výsledcích žáků utvářeny na úrovni školy, tj. mezi školami. Postup řešení modelové situace, která se zaměřuje na hodnocení vlivu rozdílů uvnitř škol a rozdílů mezi školami na celkové rozdíly ve výsledcích žáků v ověřovacím testu, je složen z následujících kroků uživatele metodiky:</p> <p>Krok 1: Uživatel metodiky vyhodnocuje výsledky žáků v ověřovacím testu, včetně volby vhodné škály pro vyhodnocení.</p> <p>Krok 2: Uživatel metodiky specifikuje a odhaduje hierarchický regresní model, kdy výsledek žáků v testu získaný v kroku 1 postupu je vysvětlován s využitím jediné proměnné odpovídající příslušnosti žáka ke škole (ID školy).</p> <p>Krok 3: Uživatel metodiky extrahuje z odhadu hierarchického regresního modelu: (a) rozptyl hodnot vztahující se k úrovni školy; a (b) rozptyl hodnot vztahující se k úrovni žáka.</p> <p>Krok 4: Uživatel metodiky počítá hodnotu vnitrotřídního koeficientu korelace (ICC) jako podíl: (a) rozptylu hodnot vztahujícího se k úrovni školy k celkovému rozptylu hodnot dosažených výsledků žáků; a (b) rozptylu hodnot vztahujícího se k úrovni žáka k celkovému rozptylu hodnot dosažených výsledků žáků.</p>
Výstup řešení modelové situace
<p>Výstupem řešení modelové situace je hodnota vnitrotřídního koeficientu korelace (ICC). Pokud je hodnota ICC rovna 0,20, pak platí, že 20 % rozptylu hodnot je vysvětleno na úrovni školy (mezi školami) a 80 % rozptylu hodnot na úrovni žáka (uvnitř škol).</p>

Širší situační kontext
<p>Modelová situace má přirozený úzký vztah k výběru vhodné škály pro vyhodnocení ověřovacího testu.</p> <p>Modelová situace je úzce zasazena do problematiky související se spravedlností ve vzdělávání, a to především v kontextu problematiky shlukování nadaných žáků ve výsledkově „lepších“ školách na jedné straně a méně nadaných žáků ve výsledkově „horších“ školách na straně druhé.</p> <p>Modelovou situaci lze rozšířit o kontrolu vlivu relevantních proměnných (např. studovaný obor žáka), a to prostřednictvím jejich zařazení do hierarchického regresního modelu jako vysvětlující proměnné.</p>
Software a ilustrace jeho využití pro řešení modelové situace
R balíček <i>ICC</i> (viz Wolak, 2016); alternativně s využitím R balíčku <i>lme4</i> (viz Bates et al., 2020)
<p><b>Postup řešení pro datový rámec TACR_DATA_HIEARCH</b></p> <pre>IDskolyf &lt;- as.factor(IDskoly)</pre> <p><i># Funkce as.factor vede k převedení dat numerické povahy na kategorická data, zde pro proměnnou označující školu.</i></p> <pre>ICCest(IDskolyf, body, data = TACR_DATA_HIEARCH)</pre> <p><i># Funkce ICCest (R balíček ICC) vede k výpočtu a zobrazení hodnoty ICC pro alternativní bodovou škálu datového rámce TACR_DATA_HIEARCH se středem 500 bodů a směrodatnou odchylkou 100 bodů.</i></p> <p><i># Atribut data označuje využívaný datový rámec pro řešení.</i></p>

### 3.3.18 Faktory vztahující se k výsledkům žáků v ověřovacím testu

Název modelové situace
Faktory vztahující se k výsledkům žáků v ověřovacím testu
Metodický postup řešení modelové situace
<p>Záměr modelové situace je spojený se zájmem uživatele metodiky nalézt ty faktory, které mají významný vztah k výsledkům žáků v ověřovacím testu. Pro správné posouzení těchto vztahů je žádoucí jejich modelování na různých úrovních hodnocení (např. žák, třída/učitel, škola), a to prostřednictvím odhadů hierarchických regresních modelů. Postup řešení modelové situace se tak skládá z následujících kroků uživatele metodiky:</p> <p>Krok 1: Uživatel metodiky vyhodnocuje výsledky žáků v ověřovacím testu na jím zvolené škále.</p> <p>Krok 2: Pro odhad hierarchických regresních modelů vybírá uživatel metodiky: (a) vysvětlovanou proměnnou v návaznosti na krok 1 postupu; (b) vysvětlující proměnné na úrovni žáka; a (c) vysvětlující proměnné na úrovni školy. Doplnit lze další úrovně hodnocení (např. třída/učitel, území).</p> <p>Krok 3: Uživatel metodiky specifikuje a odhaduje hierarchický regresní model v návaznosti na krok 2 postupu.</p> <p>Krok 4: Uživatel metodiky extrahuje odhady hierarchického regresního modelu, především pak (standardizované) hodnoty a statistickou významnost regresních koeficientů vysvětlujících proměnných.</p>

Výstup řešení modelové situace
Výstupem řešení modelové situace jsou (standardizované) hodnoty a statistická významnost regresních koeficientů vysvětlujících proměnných. Uživatel metodiky se především zajímá o faktory, které jsou statisticky významně vztaženy k výsledkům žáků v ověřovacím testu.
Širší situační kontext
Modelová situace má přirozený úzký vztah k výběru vhodné škály pro vyhodnocení ověřovacího testu. Modelová situace může mít významné praktické implikace, kdy na faktory, které jsou nejvýznamněji vztaženy k výsledkům žáků v ověřovacím testu, je možné zaměřit intervence s cílem zlepšovat vzdělávací výsledky žáků.
Software a ilustrace jeho využití pro řešení modelové situace
R balíček <i>lme4</i> (viz Bates et al., 2020); R balíček <i>lmerTest</i> (viz Kuznetsova et al., 2020); R balíček <i>sjstats</i> (viz Lüdtke, 2020)
Postup řešení pro datový rámec TACR_DATA_HIEARCH
<pre>IDskolyf &lt;- as.factor(IDskolyf)</pre> <p><i># Funkce as.factor vede k převedení dat numerické povahy na kategorická data, zde pro proměnnou označující školu.</i></p> <pre>MODEL_HIEAR &lt;- lmer(body ~ divka + vyssi_soceko_status + (1   IDskolyf), REML = FALSE, data = TACR_DATA)</pre> <p><i># Funkce lmer (R balíček lme4) vede k odhadu hierarchického regresního modelu, přičemž odhady jsou uloženy v objektu MODEL_HIEAR. Výraz (1   IDskolyf) vyjadřuje, že je zohledněna druhá úroveň hodnocení, tj. úroveň školy.</i></p> <p><i># Atribut REML označuje, že hierarchický regresní model bude odhadován metodou maximální věrohodnosti.</i></p> <p><i># Atribut data označuje využívaný datový rámec pro řešení.</i></p> <pre>summary(MODEL_HIEAR)</pre> <p><i># Funkce summary zobrazuje souhrnné výsledky odhadů hierarchického regresního modelu specifikovaného v objektu MODEL_HIEAR.</i></p> <pre>std_beta(MODEL_HIEAR)</pre> <p><i># Funkce std_beta (R balíček sjstats) vede k zobrazení standardizovaných hodnot koeficientů vysvětlujících proměnných hierarchického regresního modelu specifikovaného v objektu MODEL_HIEAR, a dále horní a dolní meze 95% intervalu spolehlivosti odhadu regresních koeficientů.</i></p>



#### 4. Inovativnost metodiky

Metodika doplňuje dosavadní přístupy v oblasti vyhodnocování výsledků ověřovacího testování v počátečním vzdělávání tím, že poskytuje komplexní metodické návody a postupy pro řešení prakticky orientovaných modelových situací. Svým komplexním pojetím integrujícím široké portfolio teoreticky opodstatněných metodických přístupů k vyhodnocování ověřovacích testů v počátečním vzdělávání a důrazem na praktickou rovinou modelových situací posunuje metodika současný stav a poskytuje nástroj s aplikačním potenciálem. Metodika dále respektuje skutečnost, že neexistuje *one-size-fits-all* řešení všech situací spojených s vyhodnocováním ověřovacího testování, a proto zachovává potřebnou flexibilitu a otevřenost vůči potenciálním uživatelům. Takto metodika poskytuje také vstupní podklady pro diskusi volby modelových situací a přístupů k jejich řešení v organizacích pracujících v oblasti ověřovacího testování v počátečním vzdělávání, čímž může dále stimulovat šíření znalostí v této problematice.

## 5. Popis uplatnění metodiky

Metodika je určena organizacím, které se ve své činnosti zabývají vyhodnocováním ověřovacích testů v počátečním vzdělávání. Takto může být metodika využita při vyhodnocování ověřovacích testů v různých kontextových situacích a pro zpracování výstupů na různých úrovních hodnocení:

- úroveň žáka (např. utváření vzdělávacích plánů ve vazbě na identifikaci silných a slabých stránek žáka);
- úroveň třídy či školy (např. výběr vhodných vzdělávacích strategií ve vazbě na identifikaci silných a slabých stránek žáků školy; využití informací pro tvorbu strategických dokumentů školy);
- úroveň systému či území (např. tvorba zpráv o stavu vzdělávacího systému České republiky v hodnocené oblasti).

Metodika zároveň umožňuje hodnotit kvalitu samotných ověřovacích testů (např. kvalita testových položek, soulad pozorovaných a modely predikovaných dat) a tím i robustnost formulovaných zjištění.

Platí přitom, že metodika poskytuje osobě, která je odpovědná za zpracování výše naznačených výstupů, podpůrný nástroj v rozhodování o způsobu jejich naplnění, a to nabídkou návodných postupů k řešení modelových situací vyhodnocení ověřovacích testů. Teprve uživatel metodiky proto rozhoduje o způsobu její aktivace a podoby uplatnění.

## 6. Seznam odkazované literatury

ALBANO, A. D. (2018). *Package 'equate'*. [online]. Dostupné z: <<https://cran.r-project.org/web/packages/equate/equate.pdf>>.

BATTAUZ, M. (2018). *Package 'equateIRT'*. [online]. Dostupné z: <<https://cran.r-project.org/web/packages/difR/difR.pdf>>.

BETEBENNER, D. (2009). Norm-and criterion-referenced student growth. *Educational Measurement: Issues and Practice*, 28(4), 42-51.

BETEBENNER, D. W., LINN, R. L. (2010). *Growth in Student Achievement: Issues of Measurement, Longitudinal Data Analysis, and Accountability*. Princeton: Educational Testing Service.

BATES, B. et al. (2020). *Package 'lme4'*. [online]. Available from: <<https://cran.r-project.org/web/packages/lme4/lme4.pdf>>.

COOK, L. L., EIGNOR, D. R. (1991). IRT equating methods. *Educational Measurement: Issues and Practice*, 10(3), 37-45.

DE CHAMPLAIN, A. F. (2010). A primer on classical test theory and item response theory for assessments in medical education. *Medical Education*, 44(1), 109-117.

DeMARS, C. (2010). *Item Response Theory*. Oxford: Oxford University Press.

EDELEN, M. O., REEVE, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research*, 16(1), 5-18.

FRYČ, J. et al. (2020). *Strategie vzdělávací politiky České republiky do roku 2030+*. Praha: Ministerstvo školství, mládeže a tělovýchovy.

CHALMERS, P. (2020). *Package 'mirt'*. [online]. Dostupné z: <<https://cran.r-project.org/web/packages/mirt/mirt.pdf>>.

KUZNETSOVA, A. et al. (2020). *Package 'lmerTest'*. [online]. Dostupné z: <<https://cran.r-project.org/web/packages/lmerTest/lmerTest.pdf>>.

LAMBERT, M. C. et al. (2018). The impact of English language learner status on screening for emotional and behavioral disorders: A differential item functioning (DIF) study. *Psychology in the Schools*, 55(3), 229-239.

- MAGIS, D., BELAND, S., RAICHE, G. (2020). *Package 'difR'*. [online]. Dostupné z: <<https://cran.r-project.org/web/packages/difR/difR.pdf>>.
- REVELLE, W. (2019b). *Package 'psych'*. [online]. Dostupné z: <<https://cran.r-project.org/web/packages/psych/psych.pdf>>.
- RIZOPOULOS, D. (2018). *Package 'ltm'*. [online]. Dostupné z: <<https://cran.r-project.org/web/packages/ltm/ltm.pdf>>.
- ROBITZSCH, A. (2020). *Package 'sirt'*. [online]. Dostupné z: <<https://cran.r-project.org/web/packages/sirt/sirt.pdf>>.
- RYAN, J., BROCKMANN, F. (2009). *A Practitioner's Introduction to Equating with Primers on Classical Test Theory and Item Response Theory*. Washington: Council of Chief State School Officers.
- TENDEIRO, J. N. (2018). *Package 'PerFit'*. [online]. Dostupné z: <<https://cran.r-project.org/web/packages/PerFit/index.html>>.
- THOMPSON, K. L. (2015). *Measuring Student Growth in K–12 Schools Using Item Response Theory within Structural Equation Models*. Hattiesburg: The University of Southern Mississippi.
- THOMPSON, N. A. (2016). *Introduction to Classical Test Theory with CITAS*. Minnetonka: Assessment Systems Corporation.
- THORPE, G. L., FAVIA, A. (2012). *Data Analysis Using Item Response Theory Methodology: an Introduction to Selected Programs and Applications*. Orono: The University of Maine.
- UDOFIA, N. A., UKO, M. P. (2016). Vertical scaling in standards-based educational assessment and accountability in educational systems. *Journal of Research & Method in Education*, 6(4), 65-75.
- VENABLES, W. N. et al. (2020). *An Introduction to R*. R Core Team. [online]. Available from: <<https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf>>.
- WILLSE, J. T. (2018). *Package 'CTT'*. [online]. Dostupné z: <<https://cran.r-project.org/web/packages/CTT/index.html>>.
- WOLAK, M. (2016). *Package 'ICC'*. [online]. Dostupné z: <<https://cran.r-project.org/web/packages/ICC/ICC.pdf>>.