Styly humoru u mužů a žen. Diferenciální fungování položek dotazníku HSQ: replikační studie[1]

Úvod a cíle

Humor Styles Questionnaire (HSQ; Martin et al., 2003) je 32položkový sebeposuzovací dotazník sestávající ze čtyř škál odpovídajících různým stylům humoru: *affiliativní* (humor pro pobavení ostatních a budování vztahů), *self-enhancing* (humor jako strategie zvládání stresu), *agresivní* (sarkasmus, zesměšňování) a *self-defeating* (sebeponižující humor za účelem sociálního přijetí). Každou dimenzi tvoří osm položek na pětibodové Likertově škále.

Muži v HSQ skórují výše než ženy, a to především na škálách agresivního a self-defeating humoru (Martin et al., 2003). Otázkou však je, zda tyto rozdíly ve skórech odrážejí skutečné rozdíly v latentním rysu sytícího danou škálu, nebo zda jsou částečně artefaktem položkového zkreslení – tedy zda některé položky fungují pro muže a ženy odlišně i při stejné úrovni sledovaného rysu. Vnímají obě skupiny respondentů položky stejně? To zjišťujeme pomocí analýzy diferenciálního fungování položek (DIF) v rámci teorie odpovědi na položku (IRT). Stejnou otázku si kladli Silvia a Rodriguez (2020), kteří vyhotovili DIF analýzu dotazníku HSQ, přičemž u souboru 1 210 respondentů zjistili nulový gender DIF. Cílem této zprávy je tento nález ověřit na nezávislém výběrovém souboru.

Dataset

Data pocházejí z veřejně dostupného datasetu[2] online administrace HSQ (pětibodová škála; 1 = *nikdy nebo velmi zřídka pravdivé* až 5 = *velmi často nebo vždy pravdivé*). Původní soubor zahrnuje 1 071 respondentů. Pro účely našeho šetření jsme vyřadili 13 osob s neuvedeným nebo nebinárním pohlavím, dále 3 osoby na základě věku, tj. mimo rozmezí 14–100 let, také jsme *list-wise* odstranili 74 respondentů s chybějící odpovědí na kterékoli položce HSQ a nakonec 10 osob s deklarovanou přesností odpovědí pod 50 %.

Výsledný analyzovaný soubor tvoří N = **971 respondentů** (533 mužů, 438 žen; věk M = 26,3, SD = 10,9).

Metoda

Diferenciálním fungováním položek (DIF, *differential item functioning*) se myslí stav, kdy se určitá položka v testu či dotazníku chová odlišně pro příslušníky různých skupin, kteří ovšem mají stejnou úroveň měřeného latentního rysu. Takto odhadujeme, zdali do fungování dané položky nevstupuje další konstrukt, který způsobuje šum v datech. Jinými slovy zjišťujeme, zdali položka měří skutečný rozdíl v míře rysu, či je výsledný rozdílný skór zapříčiněn rozličným fungováním položky pro různé skupiny, ale ve skutečnosti neodráží rozdíly v míře sledovaného rysu.

DIF lze určovat více způsoby: pomocí klasické testové teorie, ale také pomocí metod z přístupu *teorie odpovědi na položku* (IRT), která se k tomuto účelu jeví jako zvlášť vhodná (Langer et al., 2008). IRT modeluje vztah mezi latentním rysem (θ, theta) respondenta a jeho odpovědí na každou jednotlivou položku, čímž umožňuje popsat kvalitu každé položky zvlášť prostřednictvím jejích

[1] *Data a další informace o této zprávě jsou dostupné na adrese https://dostal.vyzkum-psychologie.cz/stat4?i=871*
[2] *https://openpsychometrics.org/tests/HSQ/development/*

parametrů, nikoli jen celkového skóru škály. Samotná theta θ vyjadřuje míru přítomnosti sledovaného rysu u respondenta (průměr 0, rozsah přibližně −3 až +3).

Fungování jednotlivých položek lze vizualizovat pomocí charakteristických křivek položky (ICC, item characteristic curves). Tyto křivky zobrazují pravděpodobnost volby každé odpovědní kategorie, tj. stupně 1–5 v odpovědi, v závislosti na úrovni latentního rysu θ, který jsme v rámci IRT vymodelovali. Pakliže položka funguje pro obě skupiny stejně, budou se jejich ICC překrývat (viz Obrázek 2 níže).

Rozlišujeme *uniformní DIF* (položka je systematicky snazší/těžší pro jednu skupinu: posun v obtížnosti) a *non-uniformní DIF* (položka diskriminuje různě pro obě skupiny: interakce skupiny s úrovní θ).

DIF jsme testovali pomocí balíčku *lordif* (Choi et al., 2011) v prostředí R Studia; kód jazyka R byl korigován AI modelem Claude Opus 4,6. Tento postup implementuje logistickou ordinální regresi s iterativní purifikací a porovnává tři vnořené modely: Model 1 predikuje odpověď pouze z celkového IRT skóru (baseline); Model 2 přidává hlavní efekt skupiny (test uniformního DIF); Model 3 přidává interakci skupiny se skórem (test non-uniformního DIF). Položka je označena za položku s DIF tehdy, když přidání skupinové proměnné zvýší McFaddenovo pseudo-R^2 o $\Delta R^2 \geq 0{,}01$[3]. DIF analýza byla provedena zvlášť pro každou ze čtyř škál HSQ, vždy s pohlavím jako grupovací proměnnou. Před analýzou byly reverzní položky příslušně přeskórovány.

Předpokladem IRT modelu je jednorozměrnost analyzovaných dat. V našem případě to znamená, že každá škála dotazníku by měla sytit právě jeden faktor. Tento předpoklad byl dále mimo studii od Martin et al. (2003) potvrzen na stejném datasetu v textu od Karbanové (2025).

Výsledky

Mezipohlavní rozdíly na škálách. Deskriptivní statistiky ukazují pouze malé gender rozdíly: affiliativní (Cohenovo $d = 0{,}11$), self-enhancing ($d = -0{,}04$), agresivní ($d = -0{,}11$) a self-defeating ($d = 0{,}18$), kde kladné d odpovídá vyššímu skóru mužů. Všechny rozdíly jsou pod konvenční hranicí malého efektu. Tyto efekty jsou dokonce podstatně menší než hodnoty reportované Martinem et al. (2003) a Silvií a Rodriguezem (2020), zejména na škále agresivního humoru, kde se směr rozdílu navíc obrací.

DIF analýza. Tabulka 1 shrnuje výsledky DIF analýzy pro všech 32 položek HSQ. Žádná položka nebyla označena za položku s DIF v žádné ze čtyř škál – ani na úrovni uniformního, ani non-uniformního DIF. Nejvyšší celkové ΔR^2 bylo zaznamenáno u Q15 (agresivní škála; $\Delta R^2 = 0{,}0056$) a Q19 (agresivní škála; $\Delta R^2 = 0{,}0050$); obě hodnoty jsou však hluboko pod prahem 0,01. Obrázek 1 vizualizuje celkové ΔR^2 pro všech 32 položek.

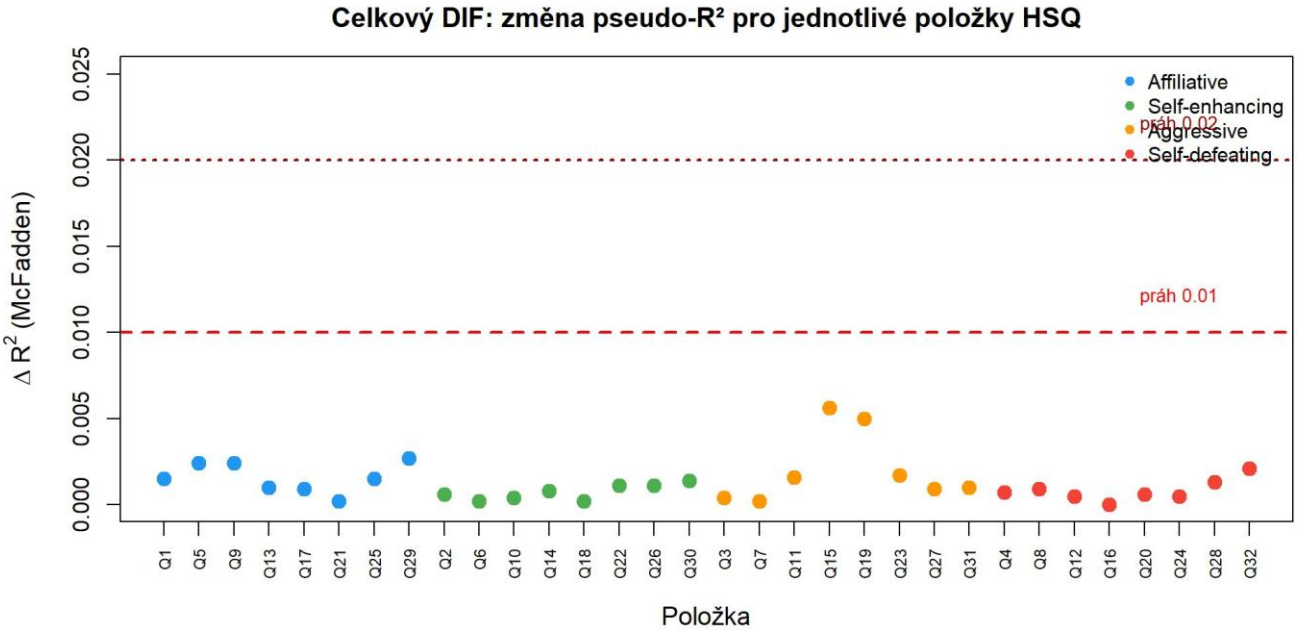
Tabulka 1. Výsledky DIF analýzy položek HSQ (McFaddenovo pseudo-R^2).

Škála	Pol.	ΔR^2 uniform	ΔR^2 non-uniform	ΔR^2 celkový	DIF
Affiliative	Q1	0,0010	0,0005	0,0015	Ne
	Q5	0,0008	0,0016	0,0024	Ne
	Q9	0,0016	0,0008	0,0024	Ne

[3]Práh $\Delta R^2 \geq 0{,}01$ je přísnější než běžně používaný 0,02 pro malý efekt; volíme jej v souladu s původní studií od Silvia a Rodrigueze (2020).

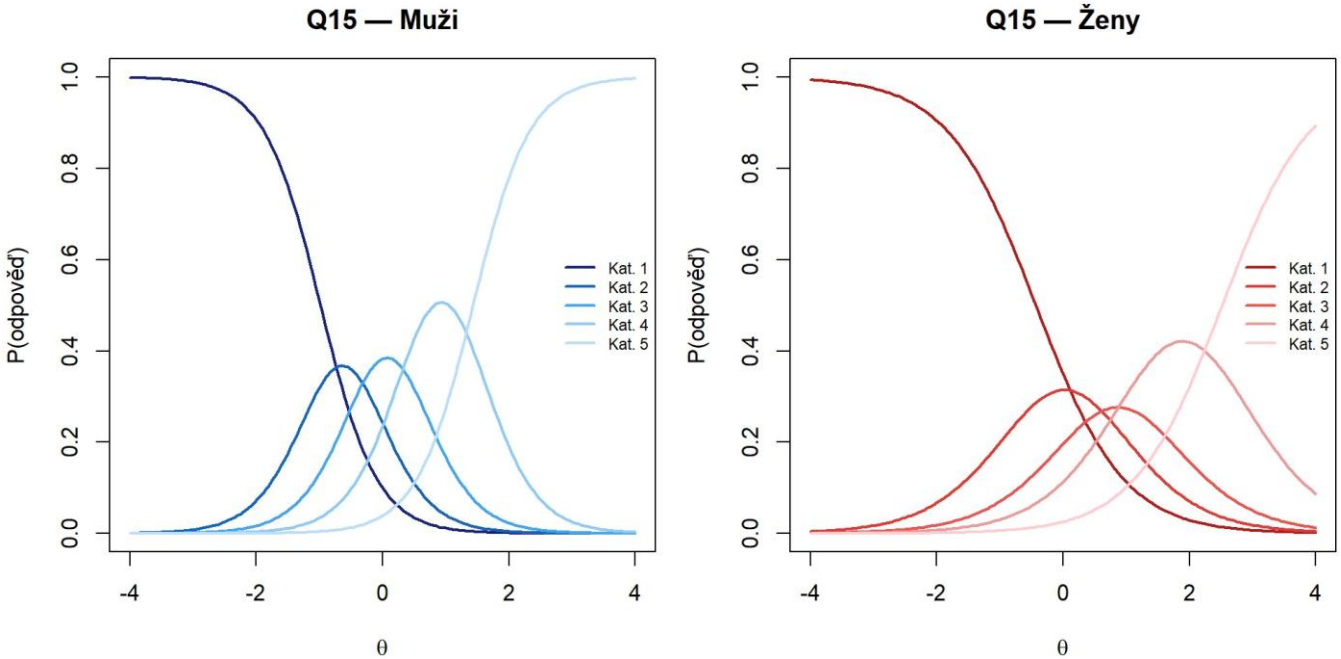
	Q13	0,0004	0,0006	0,0010	Ne
	Q17	0,0009	0,0000	0,0009	Ne
	Q21	0,0002	0,0000	0,0002	Ne
	Q25	0,0009	0,0007	0,0015	Ne
	Q29	0,0023	0,0004	0,0027	Ne
Self-enhancing	Q2	0,0002	0,0004	0,0006	Ne
	Q6	0,0002	0,0000	0,0002	Ne
	Q10	0,0004	0,0000	0,0004	Ne
	Q14	0,0001	0,0007	0,0008	Ne
	Q18	0,0002	0,0000	0,0002	Ne
	Q22	0,0010	0,0001	0,0011	Ne
	Q26	0,0001	0,0011	0,0011	Ne
	Q30	0,0014	0,0000	0,0014	Ne
Aggressive	Q3	0,0001	0,0002	0,0004	Ne
	Q7	0,0000	0,0002	0,0002	Ne
	Q11	0,0006	0,0010	0,0016	Ne
	Q15	0,0044	0,0012	0,0056	Ne
	Q19	0,0001	0,0049	0,0050	Ne
	Q23	0,0009	0,0008	0,0017	Ne
	Q27	0,0008	0,0001	0,0009	Ne
	Q31	0,0009	0,0001	0,0010	Ne
Self-defeating	Q4	0,0007	0,0000	0,0007	Ne
	Q8	0,0009	0,0000	0,0009	Ne
	Q12	0,0001	0,0004	0,0005	Ne
	Q16	0,0000	0,0000	0,0000	Ne
	Q20	0,0004	0,0002	0,0006	Ne
	Q24	0,0003	0,0002	0,0005	Ne
	Q28	0,0007	0,0006	0,0013	Ne
	Q32	0,0000	0,0021	0,0021	Ne

Poznámka: ΔR^2 uniform = změna pseudo-R^2 při přidání hlavního efektu pohlaví; ΔR^2 non-uniform = změna při přidání interakce pohlaví × skór θ (théta); ΔR^2 celkový = Model 1 vs. Model 3. Práh pro DIF: $\Delta R^2 \geq 0{,}01$.



Obrázek 1. Celkové ΔR² (McFadden) pro jednotlivé položky HSQ. Přerušovaná čára vyznačuje práh 0,01; tečkovaná čára práh 0,02. Žádná z položek nedosahuje ani nižšího prahu.

Obrázek 2 potom ilustruje charakteristické křivky odpovědi na položku (ICC) pro vybranou položku Q15 („*Nelíbí se mi, když lidé používají humor k zesměšňování druhých*“), která dosáhla nejvyššího ΔR² ze všech 32 položek, ačkoliv jde stále o velmi nízkou hodnotu. I u této položky jsou křivky pro muže a ženy prakticky zaměnitelné, což potvrzuje absenci prakticky významného DIF.



Obrázek 2. Charakteristické křivky odpovědi na položku Q15 (agresivní škála) pro muže (vlevo) a ženy (vpravo). Q15 vykazovala nejvyšší ΔR² ze všech 32 položek; i přesto jsou ICC pro obě pohlaví prakticky totožné.

Závěr

Cílem zprávy bylo ověřit, zda položky HSQ vykazují diferenciální fungování mezi muži a ženami. Výsledky jednoznačně ukazují, že žádná z 32 položek HSQ nevykazuje gender DIF – ani uniformní, ani non-uniformní – při použití McFaddenova ΔR^2 s prahem 0,01. Tím replikujeme nález Silvia a Rodrigueze (2020), kteří dospěli ke stejnému závěru na nezávislém souboru s použitím totožné metody.

Zajímavým pozorováním je, že gender rozdíly na škálách HSQ jsou v našem souboru výrazně menší než v původní validační studii (Martin et al., 2003), kde muži skórovali podstatně výše na agresivním a self-defeating humoru. V našem souboru jsou tyto efekty minimální (d = −0,11 a 0,18). To může odrážet specifika daného online výběrového souboru či posun v genderových normách mezi lety 2003 a 2012, kdy byla data naposledy aktualizována. Nulový DIF přitom naznačuje, že ať už je velikost genderových rozdílů jakákoli, **tyto rozdíly odrážejí skutečné rozdíly v latentním rysu, nikoli položkové zkreslení**.

Jako hlavní limity uvádíme samovýběr respondentů (online dotazník), nerovnoměrné věkové rozložení (medián 22 let) a absenci kontroly nad podmínkami administrace. Budoucí výzkum by mohl otestovat DIF i s ohledem na věk nebo kulturní kontext, přičemž by bylo vhodné analyzovat HSQ v překladech do dalších jazyků.

Seznam použité literatury

Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). lordif: An R package for detecting differential item functioning. *Journal of Statistical Software, 39*(8), 1–30. https://doi.org/10.18637/jss.v039.i08

Karbanová, M. (2025). *Explorativní faktorová analýza anglické verze dotazníku HSQ*. [Studentská práce do předmětu Vícerozměrné statistické metody]. https://dostal.vyzkum-psychologie.cz/stat4/zprava.php?id=392

Langer, M. M., Hill, C. D., Thissen, D., Burwinkle, T. M., Varni, J. W., & DeWalt, D. A. (2008). Item response theory detected differential item functioning between healthy and ill children in quality-of-life measures. *Journal of clinical epidemiology*, *61*(3), 268–276. https://doi.org/10.1016/j.jclinepi.2007.05.002

Martin, R. A., Puhlik-Doris, P., Larsen, G., Gray, J., & Weir, K. (2003). Individual differences in uses of humor and their relation to psychological well-being: Development of the Humor Styles Questionnaire. *Journal of Research in Personality, 37*(1), 48–75. https://doi.org/10.1016/S0092-6566(02)00534-2

Silvia, P. J., & Rodriguez, R. M. (2020). Time to renovate the Humor Styles Questionnaire? An item response theory analysis of the HSQ. *Behavioral Sciences, 10*(11), 173. https://doi.org/10.3390/bs10110173