

ve kterém nyní již umíme dosadit i za hodnotu reliability testu na základě nějakého výše uvedeného odhadu, nejčastěji Cronbachova alfa. Jak bylo uvedeno mezi předpoklady klasické teorie testů, σ_e je směrodatná odchylka normálního rozdělení chyby měření e . Odtud vyplývá i její interpretace opírající se o znalosti normálního rozdělení.

Pokud žák dosáhne v testu výsledek t , pak jeho skutečná dovednost T se nachází v intervalu $(t - e, t + e)$ s pravděpodobností přibližně 68 %, při rozšíření tohoto intervalu na $2e$ již máme jistotu 95 %. V intervalu $(t - 3e, t + 3e)$ se žákova dovednost T již nachází s pravděpodobností hraničící s jistotou 99,7 %.

Příklad 4.64 Chyba měření odvozená z reliability testu

Reliabilita testu z příkladu 4.62 je 0,944. Směrodatná odchylka výsledků žáků v daném testu o 61 úlohách byla 13,6. Standardní chyba měření pak vychází 3,2.

Žákova „skutečná“ znalost a dovednost v anglickém jazyce (vyjma písemného a ústního projevu) se s 95% pravděpodobností nachází v intervalu plus minus 6,4 bodu kolem aktuálně žákem dosaženého výsledku v testu. Téměř se 100% jistotou se nachází v intervalu plus minus 9,6 bodu kolem aktuálně žákem dosaženého výsledku v testu.

Zdroj: NIQES 2014, test z anglického jazyka pro 8. ročník, varianta 1, řešilo 3 558 žáků.

Z předešlého textu ale též víme, že reliability je možné odhadovat jen některými empirickými cestami sběru dat a následných výpočtů. Odsud ale plyne, že určitý způsob odhadu reliability může odhalit jen určitou část chyby měření, která se při daném postupu může projevit a vede např. při ekvivalenci k tomu, proč žák dosáhl jiného výsledku ve verzi testu A a verzi testu B. V následující tabulce ve schématu 4.7 jsou shrnuty možné zdroje chyb a odpovídající postupy odhadu reliability.

Schéma 4.7 Souvislost mezi zdrojem náhodné chyby měření a postupem odhadu reliability

Zdroj náhodné chyby	Postup odhadu reliability
„Rozpoložení“ osoby v čase testu	Dependabilita (test-retestová reliability)
Specifičnost výběru položek do verze testu, včetně „rozpoložení“ osoby v čase testu	Ekvivalence
Specifičnost každé položky v testu vzhledem ke zjišťované dovednosti	Vnitřní konzistence

Vedle uvedených postupů odhadu reliability, které byly pojednány v předcházející části, lze přístup uplatnit i na některé specifické případy. Takovým specifickým případem může být expertní posuzování. Můžeme si jej představit při využívání hodnoticí škály nebo škál při posuzování různých žákovských výkonů. Zjednodušíme si to na případ, kdy hodnotitelé hodnotí otevřené úlohy v testu. Jak již víme, míra otevřenosti úlohy může být různá. Jsou úlohy doplňovací, kde míra subjektivity při hodnocení je minimální, ale pak jsou úlohy, kde míra subjektivity hodnotitele je větší. Subjektivita hodnotitele je pak zdrojem chyby, která se promítá do nepřesnosti měření zjišťované dovednosti v testu. Ponechme nyní stranou cesty snižování subjektivity hodnocení jako roli třetího hodnotitele v případě neshod a podobně. Uvažme následující situaci.

Máme dva (nebo více) stejně kvalitních hodnotitelů. Výsledek hodnocení v dané úloze je dán rozhodnutím hodnocení jednoho z nich (nejsou zde tedy mechanismy kontroly a zpřesňování