

O oprávnění interpretovat výsledky za určitou část testu lze uvažovat teprve tehdy, pokud

$$r_M > r_{M/N}$$

Pro posuzování psychometrické relevance prezentování výsledků za dílčí části testů (subtesty, obsahové oblasti, oblasti dovedností, viz oddíl 4.1.4) se dá použít porovnání reliability spočtené z úloh v dané části s odhadem reliability náhodné skladby úloh z celého testu stejně početně zastoupené, jako má analyzovaná část (podrobněji k tomu viz výše). Samozřejmě čím je reliability spočtená z úloh v dané části testu vyšší než reliability náhodné skladby úloh v testu, tím je interpretace výsledku za danou část testu oprávněnější. Viz příklad 4.63 (pro tento příklad byla použita stejná data jako v příkladu 4.61, hodnoty v tabulce je možno odečíst i z výše uvedeného grafu), kde v případě částí slovní zásoba a poslech toto platí, ale pro část gramatika nikoliv. Část gramatika se skládá z takových úloh, jejichž vnitřní konzistence je menší než náhodná skladba 18 úloh vybraných z celého testu, a výsledek za tuto část tedy postrádá interpretační význam jako výsledek o gramatických znalostech a dovednostech žáků. Může to být dáno i tím, že z obsahového hlediska je oblast gramatiky velmi široká. Naopak výsledky za části týkající se znalostí slovní zásoby a poslechových dovedností lze interpretovat relativně spolehlivě.

#### Příklad 4.63 Reliability dílčích částí testu

subtest	počet úloh	reliabilita	odvozená reliability z celého testu	rozdíl
Čtení s porozuměním	15	0,818	0,804	0,013
Gramatika	18	0,806	0,832	-0,025
Slovní zásoba	18	0,885	0,832	0,053
Poslech	10	0,773	0,733	0,040
Celý test	61	0,944	0,944	

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

#### Závislost reliability na vzorku testovaných žáků

Při nahlížení na koeficient reliability jako na parametr pro posuzování kvality testu je potřeba si uvědomit jeho důležitou vlastnost, a sice závislost na různosti dovedností žáků, kterým byl test zadán, pro odhad reliability. Ze vzorce, který reliability teoreticky definuje, je vidět, že čím bude rozptýl testových výsledků větší (větší různost dovedností testovaných žáků), tím bude reliability vyšší. Proto je přirozené, že reliability na různých testovaných vzorcích žáků může vycházet trochu jinak. Stabilně by však měla již vycházet chyba měření uvedená výše.

Pokud se tedy např. spočítá reliability testu zvlášť na vzorcích žáků z víceletého gymnázia a druhého stupně základních škol, je víceméně jisté, že tyto odhady reliability budou nižší, než reliability testu odhadnutá z celé žakovské populace.

Podobně tomu je i u společné části maturitní zkoušky. Reliability testů zvlášť na gymnáziích, středních odborných školách i středních odborných učilištích by byly nižší než na celém vzorku všech maturujících žáků. Též není možné porovnávat reliability testů např. z českého jazyka, který řeší u maturity všichni žáci, a reliability z výběrových zkoušek. Stejně tak je nesrovnatelná reliability testu mezi jarním a podzimním termínem maturitní zkoušky.