

TIPY A RADY PŘI SESTAVOVÁNÍ TESTOVÝCH ÚLOH Z CHEMIE

MILADA TEPLÁ a MARTIN ŠRÁMEK

Katedra učitelství a didaktiky chemie, Přírodovědecká fakulta Univerzity Karlovy, Albertov 6, 128 00 Praha 2, Česká republika
milada.tepla@natur.cuni.cz

Došlo 20.1.23, přijato 17.3.23.

Článek prezentuje tipy a doporučení pro tvůrce testových úloh nejen z chemie. Nejprve je vymezena základní charakteristika testů, testových položek i samotných distraktorů. Následně je představeno sedm doporučení na konkrétních dvanácti příkladech testových položek, pro které byla vypočtena úspěšnost řešitele, citlivost a též provedena analýza distraktorů. Taktéž bylo poukázáno na nedostatečnou korelaci mezi reálnou a očekávanou obtížností testových položek.

Klíčová slova: položková analýza, citlivost, obtížnost, atraktivita distraktoru, kvalita testu

1. Testové úlohy ve výuce chemie

Jedním z nástrojů, který slouží ke zjišťování vědomostí a dovedností žáků/studentů, je znalostní test¹. Pedagog při přípravě testů si musí být vědom, co má prostřednictvím testu zjišťovat a na základě toho volit vhodné testové položky – jak z obsahového, tak technického hlediska. Z hlediska odborné terminologie je možné rozlišit několik druhů testů. Jedním z druhů může být tzv. didaktický test, který slouží jako nástroj k systematickému měření výsledků výuky¹. Dalším typem je přijímací test sloužící jako nástroj k výběru kvalitních uchazečů pro studium na vysoké škole.

Aby byl vzniklý test testem kvalitním a zkoumal to, k čemu byl skutečně sestaven, je nutné věnovat velikou pozornost přípravě testů a jejich vyhodnocení.

2. Jaké parametry určují kvalitu testu, vhodnost zařazení úloh a správnost volby distraktorů?

2.1. Kvalita testu

Mezi důležité otázky při posuzování kvality testů patří, zda použitý test byl dostatečně spolehlivý a přesný a zároveň testoval ty znalosti a poznatky, které testovat měl. Z toho důvodu se zkoumají dva významné parametry didaktického testu: reliabilita (spolehlivost a přesnost) a obsahová validita (obsahová správnost).

Reliabilita vyjadřuje, zda byl test z technického hlediska sestaven kvalitně (spolehlivě a přesně)², tedy zda bychom při opakovaném použití získali obdobné výsledky. Pro odhad reliability je možné použít koeficient Cronbachova alfa, který vyjadřuje vnitřní konzistenci testu a nabývá hodnot 0 až 1 s tím, že pro didaktické testy bývá, má-li být

test považován za konzistentní, dle literatury požadována hodnota Cronbachova alfa alespoň 0,7 (cit.³).

Validita znamená schopnost výzkumného nástroje zjišťovat to, co zjišťovat má⁴. V tomto případě schopnost didaktického testu ověřovat požadované znalosti. Obsahová validita je definována jako míra pokrytí oblasti vědomostí a poznatků, které mají být testovány⁵.

Obsahová validita testů může být posuzována expertním panelem, který se může skládat z odborníků, didaktiků a též učitelů. Mezi panelisty mohou být začleněny dále tvůrce testů a kontrolori, kteří testy po jejich vytvoření kontrolují a navrhuji jejich úpravy. Panelisté mj. posuzují, zda jsou testové položky jednoznačně formulované, pro uchazeče srozumitelné a zda umožňují výběr jednoznačné odpovědi.

U přijímacích testů je velmi významná jeho predikční validita, která vypovídá o tom, do jaké míry je test schopen předpovídat výsledky studenta v budoucnu v oblastech, které nás primárně zajímají a kvůli kterým je test vytvořen⁵.

2.2. Vhodnost úloh (testových položek)

Pro zjištění, zda vytvořené úlohy byly vhodné pro jejich zařazení do testu, se jako významné parametry uvádí obtížnost úlohy (popř. úspěšnost řešitele) a její diskriminační schopnost (citlivost).

Jak je úloha obtížná, je vypočteno prostřednictvím úspěšnosti řešitelů v dané úloze, která je definována jako podíl řešitelů, kteří úspěšně vyřešili danou úlohu, a celkového počtu řešitelů⁵. Úspěšnost řešitelů v úloze nabývá hodnot 0 až 1 s tím, že v didaktických testech jsou doporučovány úlohy s obtížností, resp. úspěšností v rozmezí 0,2 až 0,8 (cit.¹). U uzavřených úloh, při kterých řešitel vybírá právě jednu správnou odpověď ze čtyř alternativ, bývá jako *obtížná úloha* považována úloha s úspěšností řešitele nižší než 0,3 (cit.⁶).

Citlivost úlohy (neboli její diskriminační schopnost) vyjadřuje, zda úspěšnější řešitelé v celém testu byli též úspěšnější i v analyzované úloze. Zda tomu tak skutečně bylo, tedy zda úloha dostatečně rozlišovala úspěšné a neúspěšné řešitele neboli byla dostatečně citlivá (z hlediska konstrukce testu „efektivní“), je posuzováno na základě např. korelačního koeficientu RIR či koeficientu citlivosti ULI.

Korelační koeficient RIR je korelační koeficient mezi bodovým ziskem každého uchazeče v dané položce a jeho celkovým bodovým ziskem v testu bez započtení bodů za danou položku⁵. Za *vhodně citlivou úlohu* je považována taková, jejíž hodnota koeficientu RIR je alespoň 0,15. Velmi dobré úlohy vykazují koeficient RIR vyšší než 0,25 (cit.⁷).

Pro výpočet koeficientu citlivosti $ULI(\frac{1}{2})$ jsou všichni řešitelé seřazeni dle celkové úspěšnosti dosažené v daném testu a následně rozděleni na dvě poloviny. Pro každou skupinu je následně vypočtena úspěšnost dané úlohy. Rozdíl těchto hodnot je roven právě koeficientu $ULI(\frac{1}{2})$ ⁵. Citlivost bývá ve spojitosti s jejich úspěšností odstupňovaná, neboť u úloh s úspěšností do 30 % (či naopak u úloh s úspěšností velmi vysokou) bývá v literatuře snížen požadavek na koeficienty citlivosti ($ULI(\frac{1}{2})$ a RIR) s ohledem na jejich zvýšenou náročnost¹⁷. Za úlohu s *dostatečnou citlivostí* (dle $ULI(\frac{1}{2})$) lze považovat takovou testovou otázku, která splňuje jednu z podmínek¹: a) Obtížnost úlohy je nižší než 0,3 a zároveň hodnota vypočteného koeficientu $ULI(\frac{1}{2})$ je vyšší nebo rovna 0,15; b) Obtížnost úlohy je mezi 0,3 až 0,7 a zároveň hodnota vypočteného koeficientu $ULI(\frac{1}{2})$ je vyšší nebo rovna 0,25; c) Obtížnost úlohy je vyšší než 0,8 a zároveň hodnota vypočteného koeficientu $ULI(\frac{1}{2})$ je vyšší nebo rovna 0,15.

2.3. Vhodnost distraktorů

U uzavřených testových položek s výběrem jediné správné odpovědi ze 4 alternativ je též velmi podstatné provést analýzu navržených distraktorů (= chybné alternativy). Zjišťuje se, kolik procent řešitelů distraktor volilo (atraktivita distraktoru) a zda byl distraktor volen lepšími či slabšími řešiteli (diskriminace)⁵.

Co se týče atraktivity distraktorů, lze definovat následující pojmy: *nefunkční distraktor*, což je distraktor, který volí méně než 2 % testovaných, a *neatraktivní distraktor*, což je distraktor, který volí 2 až 5 % testovaných⁹.

Diskriminace distraktoru se vypočítává obdobně jako $ULI(\frac{1}{2})$, ale ne pro správnou odpověď, ale pro jednotlivé distraktory (rozdíl úspěšností v dané úloze mezi žáky s nejlepšími a nejllepšími výsledky)⁵, tj. zda tuto (chybnou) odpověď volili spíše studenti, kteří získali v testu celkově podprůměrný počet bodů, či naopak. Distraktor s kladnou diskriminací je *nehodný*, neboť kladná diskriminace distraktoru svědčí o skutečnosti, že tuto nesprávnou odpověď volili spíše řešitelé, kteří celkově v porovnání s ostatními dosáhli lepších výsledků, a proto je vhodné zvážit úpravu tohoto distraktoru.

3. Doporučení při sestavování testových úloh – příklady úloh a jejich vlastností

Testové úlohy by měly být zcela jednoznačné a pro danou věkovou skupinu řešitelů zcela srozumitelné. Nad to by tvůrci testů měli dbát následujících doporučení pro tvorbu testů⁸: 1) vyhněte se negativně formulovaným úlohám; 2) nepoužívejte relativizující ani absolutní termíny; 3) nevytvářejte distraktory, které řešitel logicky vyloučí, aniž by musel úlohu řešit („opakující je správně“); 4) zaměřte se na významné problémy, nezařazujte do testů chytáky; 5) nezařazujte odpovědi typu „všechny uvedené odpovědi“ či „nic z uvedeného“. Dále autoři článku doporučují: 6) aby úlohy neobsahovaly redundantní části, tj. části zadání, které nejsou při řešení úloh potřeba a 7) u výpočtových úloh neuvádět číselné hodnoty, ke kterým nelze dospět za využití nejpravděpodobněji uvažovaných matematických operací („nesmyslné hodnoty“).

V následujících podkapitolách jsou uvedeny příklady úloh, které porušovaly jedno z výše uvedených doporučení. U každé úlohy byla provedena analýza úlohy a též analýza distraktorů. Úlohy byly součástí oborových testů z chemie, které byly použity mezi lety 2016 až 2019 v rámci přijímacího řízení na Přírodovědeckou fakultu Univerzity Karlovy. Každý test byl řešen minimálně 213 a maximálně 238 uchazeči. Řešitelé vybírali vždy jednu správnou odpověď ze čtyř alternativ s tím, že za nesprávnou odpověď nebyly uděleny záporné body. Úlohy byly bodovány v rozmezí 2 až 6 bodů dle jejich předpokládané náročnosti. Reliabilita testů se pohybovala na přijatelné úrovni v rozmezí 0,66 až 0,78. Celkem bylo analyzováno 240 testových položek, z nichž většina dosahovala přijatelné citlivosti (citlivost 91 % položek dosahovala alespoň u jednoho z koeficientů $ULI(\frac{1}{2})$ či RIR požadované hodnoty). Obsahovou validitu řešil expertní panel skládající se z pěti panelistů (3 tvůrci testu a 2 kontrolori). Taktéž byla zkoumána predikční validita ve vztahu k úspěšnému dokončení bakalářského studia⁹.

3.1. Negativně formulované úlohy

Úloha zaměřená na karbidy (obr. 1) je příkladem negativně formulované úlohy. Úloha vykazuje přibližně průměrnou úspěšnost (53 %), avšak nízkou citlivost – $ULI(\frac{1}{2}) = 0,12$; RIR = 0,17. Dokonce distraktor a) je nevhodným distraktorem. Domníváme se, že je tomu právě z důvodu, že část uchazečů si nevěšme negativní formulace zadání a ve spěchu a stresu z testování označí první pravdivé tvrzení.

Druhým příkladem negativně formulované úlohy je úloha, která je zaměřená na hydrolyzu proteinů (obr. 2). Úspěšnost v úloze činila 43 %. Na druhou stranu tato úloha vykazuje dokonce zápornou citlivost – koeficient $ULI(\frac{1}{2})$ je roven hodnotě $-0,07$, koeficient RIR hodnotě $-0,14$. Na základě uvedených hodnot lze tvrdit, že úloha nepřispívala k výběru vhodných řešitelů. Úloha obsahuje 2 atraktivní, ale zároveň nevhodné distraktory b) a c) v kombinaci s jedním neatraktivním distraktorem a). Níz-

Vyberte nesprávné tvrzení.

- a) Karbid vápníku má vzorec CaC_2 .
b) Reakcí Al_4C_3 s vodou vzniká kyselina hlinitá.
 c) Hydrolýzou CaC_2 vzniká acetylen.
 d) při hydrolýze acetylidu vápenatého vzniká též hydroxid vápenatý.

Obr. 1. **Negativně formulovaná úloha** (úsp. 53 %, $\text{ULI}(\frac{1}{2}) = 0,12$, $\text{RIR} = 0,17$; a) je nevhodný distraktor).

Vyberte sloučeninu či směs sloučenin, kterou nemůžeme nalézt mezi produkty hydrolýzy běžných proteinů:

- a) aminokyseliny b) fosfáty c) sacharidy **d) sírany**

Obr. 2. **Negativně formulovaná úloha** (úsp. 43 %, $\text{ULI}(\frac{1}{2}) = -0,07$, $\text{RIR} = -0,14$; b) a c) jsou nevhodné distraktory, a) je neatraktivní distraktor).

ká atraktivita distraktoru a) může být důsledkem toho, že většina uchazečů ví, že proteiny se skládají z aminokyselin – jedná se o základní znalost vyučovanou na SŠ. Řešitelé pak volili mezi možnostmi b), c) a d) s tím, že v testu úspěšnější volili spíše mezi distraktory b) a c), což mohlo být následkem negativně formulované úlohy.

3.2. Používání relativizujících nebo absolutních termínů

Nepoužívejte relativizující (např. většinou, možná, obvykle) ani absolutní termíny (např. všechny, nikdy, musí). Příkladem takové úlohy je úloha zaměřená na hydrataci uhlovodíku (viz obr. 3). Správnou odpovědí je alternativa c), tedy jediná, u níž není použit absolutní termín „pouze“. Nicméně i přes tuto výhradu je úspěšnost řešitelů v úloze pouze 42,9 % a citlivost dosahuje velmi dobrých hodnot – koeficient $\text{ULI}(\frac{1}{2})$ nabývá hodnoty 0,24, koefi-

cient RIR 0,17. Úloha tedy přispěla k výběru nejlepších řešitelů.

3.3. „Opakující je správně“

Při tvorbě testu je velmi vhodné vyhnout se tomu, aby nejčtenější odpověď uvedená u jednotlivých distraktorů byla správná. Příkladem takové úlohy je úloha zaměřená na praktické využití organických látek (obr. 4). Úspěšnost řešitelů byla 82 %, tedy lze ji považovat za „příliš snadnou“. Koeficient $\text{ULI}(\frac{1}{2})$ nabývá hodnoty 0,17, koeficient RIR 0,20, což poukazuje na relativně vysokou citlivost dané úlohy vzhledem k velmi vysoké úspěšnosti. Z nabízených distraktorů jsou však dva neatraktivní (odpovědi b) a c)), což mohlo být způsobeno právě frekvencí kombinací „A4“ a „B3“ v odpovědích, které vylučují distraktory b) a c). Uchazeči tedy volili mezi možnostmi a) a d).

Hydratací acyklického uhlovodíku o souhrnném vzorci C_4H_6 vzniká butanon. Acyklický uhlovodík může být.

- a) pouze but-1-yn b) pouze buta-1,3-dien **c) but-1-yn nebo but-2-yn** d) pouze but-2-yn

Obr. 3. **Úloha s absolutními termíny** (úsp. 42,9 %, $\text{ULI}(\frac{1}{2}) = 0,24$, $\text{RIR} = 0,17$).

Přiřad'te k sobě správně sloučeninu a oblast jejího využití v praxi.

- | | |
|--------------------------------|-------------------------|
| (A) isooktan | (1) sváření kovů |
| (B) kyselina salicylová | (2) plasty |
| (C) tetrafluorethylen | (3) léčiva |
| (D) acetylen | (4) motorismus |

Správné řešení je:

- a) **A4, B3, C2, D1**
 b) A1, B3, C4, D2
 c) A4, B2, C1, D3
 d) A4, B3, C2, D4

Obr. 4. **„Opakující se je správně“** (úsp. 82 %, $\text{ULI}(\frac{1}{2}) = 0,17$, $\text{RIR} = 0,20$; b) a c) jsou neatraktivní distraktory).

3.4. Zaměříme se na podstatné

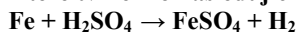
Tvůrci testů by se měli zaměřit na významné problémy a do testů nezařazovat chytáky. Příkladem velmi náročné úlohy je úloha zaměřená na tvrzení týkající se reakce železa s kyselinou sírovou (obr. 5). Úspěšnost této úlohy činila pouze 25 %. Poslední nabízená alternativa (a zároveň správná odpověď d) se zaměřovala na objem vznikajícího vodíku za laboratorních podmínek. Lze se domnívat, že náročnost úlohy souvisí s tím, že na středních školách v rámci výuky chemie není vyučována stavová rovnice ideálního plynu, kterou by bylo možné využít k ověření, že odpověď d) je správná. Úspěšnost v této úloze je dle našeho názoru možné zvýšit úpravou alternativy d), ve které by žáci určovali objem vznikajícího plynu při 0 °C místo za podmínek laboratorních. Přes velmi nízkou úspěšnost však byla tato úloha velmi citlivá ($ULI(\frac{1}{2}) = 0,30$, $RIR = 0,36$), čímž přispěla k výběru nejlepších řešitelů. Zajímavé je srovnání této úlohy s úlohou zaměřenou na reakci mědi s kyselinou sírovou (obr. 6). Úspěšnost v této na první pohled velmi podobné úloze byla však velmi odlišná (65 %).

Další úlohou je úloha zaměřená na vlastnosti fluorovodíku (obr. 7). Úspěšnost řešitelů byla 33 %, což poukazuje na poměrně vysokou náročnost úlohy. Koeficient $ULI(\frac{1}{2})$ u této úlohy dosahuje hodnoty pouze 0,04, koeficient RIR

0,00, což svědčí o její velmi nízké citlivosti. Navíc odpovědi c) a d) jsou nevhodnými distraktory, tedy jsou častěji zvoleny řešiteli patřící mezi polovinu úspěšnějších řešitelů. Možnou příčinou nízké úspěšnosti v této úloze je skutečnost, že při výuce na středních školách nebývá příliš věnována pozornost komparaci teplot varu hydridů v rámci jedné skupiny prvků navzdory skutečnosti, že tato problematika je detailně představena v některých středoškolských učebnicích¹⁰. S ohledem na výše uvedená fakta lze konstatovat, že úloha není dostatečně selektivní.

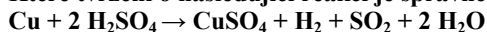
3.5. „Všechny uvedené odpovědi“ či „nic z uvedeného“

Dle literatury⁸ po začlenění alternativy typu „všechny uvedené odpovědi“ či „nic z uvedeného“ mezi distraktory, pak řešitelé tuto možnost upřednostňují. To se projevilo též u úlohy zaměřené na reakci 2-bromopropanu s ethoxidem (obr. 8). Úspěšnost v této úloze je velmi nízká – 27 %, citlivost je adekvátní – $ULI(\frac{1}{2}) = 0,21$, $RIR = 0,20$, avšak odpověď d) je nevhodným distraktorem, který volilo celkem 34,2 % všech řešitelů a vzhledem k jeho záporné diskriminanci by měl být nahrazen jiným distraktorem, neboť nepřispíval k selekci úspěšných řešitelů. Navíc daná reakce může probíhat jak mechanismem eliminačním, tak substitučním.

Které tvrzení o následující reakci je správné?

- Kyselina sírová má v této reakci redukční vlastnosti.
- Reakce probíhá pouze v koncentrované kyselině sírové.
- Vzniklá sůl je nerozpustná ve vodě a vzniká tedy sraženina.
- Ze 2 molů železa vzniká za laboratorních podmínek asi 48 dm³ vodíku.**

Obr. 5. Zaměříme se na podstatné (úsp. 25 %, $ULI(\frac{1}{2}) = 0,30$, $RIR = 0,36$).

Které tvrzení o následující reakci je správné?

- Reakce probíhá pouze se zředěnou kyselinou sírovou.
- Z jednoho molu kyseliny sírové vzniká jeden mol oxidu siřičitého.
- V systému vzniká sraženina síranu měďného.
- Kyselina sírová má vlastnosti oxidačního činidla.**

Obr. 6. Zaměříme se na podstatné – srovnání (úsp. 65 %, $ULI(\frac{1}{2}) = 0,18$, $RIR = 0,26$).

Vyberte správné tvrzení o fluorovodíku:

- má vyšší teplotu varu než chlorovodík**
- je to silná kyselina
- ve vodě zcela disociuje
- v reakcích vykazuje oxidační účinky

Obr. 7. Zaměříme se na podstatné (úsp. 33 %, $ULI(\frac{1}{2}) = 0,04$, $RIR = 0,00$; c) a d) jsou nevhodnými distraktory).

Reakcí 2-bromopropanu s ethoxidem (ethanolátem) sodným vznikají dva produkty:

- a) ethyl(isopropyl)ether a propen
 b) propen a bromethan
 c) propanol a ethyl(isopropyl)ether
 d) reakce neprobíhá

Obr. 8. „Nic z uvedeného“ (úsp. 27 %, $ULI(\frac{1}{2}) = 0,21$, $RIR = 0,20$; d) je nevhodným distraktorem).

3.6. Redundantní části

Doporučujeme, aby úlohy neobsahovaly redundantní části, tj. části zadání, které nejsou při řešení úloh potřeba. Zadání takovéto úlohy zaměřené na aromatický charakter látek je na obr. 9. Řešitelé mají na výběr 8 sloučenin, z nichž 2 se však v odpovědích vůbec nevyskytují, a proto je zbytečné je do zadání uvádět. Úspěšnost řešitelů v této úloze dosahovala hodnoty 90 %, koeficient $ULI(\frac{1}{2})$ byl roven hodnotě 0,21, koeficient RIR hodnotě 0,37, což svědčí o poměrně velmi vysoké citlivosti dané úlohy (úloha tedy selektovala nejméně úspěšné řešitele od ostatních), avšak odpověď c) je neatraktivním distraktorem a odpověď d) dokonce nefunkčním distraktorem. Příčinou nízké atraktivity těchto odpovědí může být skutečnost, že většina uchazečů ví, že fenol je aromatická sloučenina, a proto varianty c) a d) snadno vyloučí. Vhodnou úpravou by bylo, že by řešitelé nevybírali všechny aromatické látky z kompletní nabídky sloučenin, ale vybírali by alternativy, ve kterých jsou uvedeny pouze aromatické látky.

3.7. „Nesmyslné hodnoty“

Obsahují-li distraktory u výpočtových úloh hodnoty, k nimž řešitel nemůže dospět na základě pro výpočet uvažovaných matematických operací, přispívá to k jejich nízké atraktivitě. Tento jev lze pozorovat u úlohy zaměřené

na výpočet počtu nukleotidů kódující daný počet aminokyselin (obr. 10). Úspěšnost řešitelů byla v této úloze 74 %, citlivost $ULI(\frac{1}{2}) = 0,14$, $RIR = 0,08$ (tj. jedná se o necitlivou úlohu). Distraktor d) byl vyhodnocen jako neatraktivní, nabízenou hodnotu řešitelé vyhodnotili jako nesmyslnou. Závěry lze podpořit i při srovnání s velmi obdobnou úlohou (obr. 11), která vykazuje zcela odlišné vlastnosti: úspěšnost 55 % a poměrně vyšší koeficienty citlivosti – $ULI(\frac{1}{2}) = 0,24$, $RIR = 0,17$. Z porovnání dat z těchto dvou úloh lze usuzovat, že část uchazečů si není jistá, zda poměr počtu bází RNA ku počtu kódovaných aminokyselin je roven poměru 3 : 1 či převrácenému poměru, tj. 1 : 3. Tato hypotéza vysvětluje poměrně vysokou atraktivitu distraktorů a) (volen 18 % uchazečů) a b) (volen 7 % uchazečů) v druhé uvedené otázce. Jelikož v první uvedené otázce není možné poměr 3 : 1, resp. 1 : 3 zaměnit, je možné snadněji dedukovat správnou odpověď, což vede ke zvýšení úspěšnosti a snížení citlivosti úlohy. Závěry podporují hypotézu zdůvodňující, proč v první úloze byl distraktor d) vyhodnocen jako neatraktivní. Ve výpočtu by musel být použit poměr 9 : 1, což řešitelé v téměř 98 % vyloučili.

Dalším příkladem úlohy s distraktory obsahující „nesmyslné hodnoty“ je úloha zobrazená na obr. 12. Úspěšnost této úlohy činila 80,2 %; citlivost pak $ULI(\frac{1}{2}) = 0,25$; $RIR = 0,27$, avšak obsahovala dva neatraktivní distraktory – b) a c).

Které z uvedených sloučenin mají aromatický charakter?

- | | | | |
|----------------------|-----------------------|----------------------------|---------------------------|
| A) fenol | B) cyklopenten | C) kyselina ftalová | D) cyklohexadien |
| E) cyklohexan | F) fenantren | G) cyklopropen | H) cykloheptatrien |
- a) A, C, F
 b) A, E, F
 c) E, G, H
 d) C, G, H

Obr. 9. **Redundantní části** (úsp. 90 %, $ULI(\frac{1}{2}) = 0,21$, $RIR = 0,37$; c) je neatraktivním distraktorem a d) nefunkčním distraktorem).

Jaký minimální počet nukleotidů musí mít ribonukleová kyselina, aby mohla kódovat protein složený z 200 aminokyselin?

- a) 203 b) 197 c) 603 d) 1 803

Obr. 10. „Nesmyslné hodnoty“ (úsp. 74 %, $ULI(\frac{1}{2}) = 0,14$, $RIR = 0,08$; d) je neatraktivním distraktorem).

Jaký minimální počet nukleotidů musí mít ribonukleová kyselina, aby mohla kódovat protein složený z 600 aminokyselin?

- a) 203 b) 197 c) 603 **d) 1 803**

Obr. 11. „Nesmyslné hodnoty“ – srovnání (úsp. 55 %, ULI(½) = 0,24, RIR = 0,17).

Kolik molů jodidu olovnatého vznikne při reakci 5 molů jodidu draselného s nadbytkem dusičnanu olovnatého?

- a) 5 b) 3,5 c) 3 **d) 2,5**

Obr. 12. „Nesmyslné hodnoty“ (úsp. 80,2 %, ULI(½) = 0,25, RIR = 0,27; b) a c) jsou neatraktivními distraktory).

4. Další závěry plynoucí z položkové analýzy

4.1. Doporučení pro bodové ohodnocení úloh

Rozbor osmi oborových testů z chemie ukázal poměrně nízkou závislost mezi bodovým ohodnocením úlohy (autory testu zamýšlenou obtížností) a její skutečnou (vypočtenou) obtížností (resp. úspěšností řešitele). Korelační koeficient se pohyboval v rozmezí 0,06 až 0,35 s tím, že průměrná hodnota byla 0,14. Taktéž bylo zjištěno, že se vlastnosti jednotlivých úloh (např. úspěšnost, citlivost) výrazně liší v závislosti na pořadí nabízených odpovědí nebo pořadí položky v testu. Z tohoto důvodu pravděpodobně neexistuje jednoznačný model, který by mohl být pro přiřazení adekvátního bodového ohodnocení jednotlivým testovým položkám použit. Při ohodnocení všech úloh shodným počtem bodů se zpětným výpočtem zjistilo, že vždy dochází k velmi mírnému nárůstu koeficientů reliability na hodnoty 0,69 až 0,82. Z výše uvedených důvodů autoři testu doporučují všem testovým položkám přiřadit shodný počet bodů.

4.2. Příčiny snížené citlivosti položky

Přestože se lze domnívat, že citlivost úlohy silně závisí na konkrétním zadání, na základě realizovaného rozboru celkem 240 testových položek lze nalézt několik příčin snížené citlivosti: (a) vysoký počet kroků nutných k vyřešení úlohy (např. u slovně zadané redoxní reakce); (b) příliš lehké nebo naopak příliš těžké úlohy, u kterých uchazeči svoji odpověď pouze tipovali, či odvozovali, aniž by využili chemické znalosti; (c) přehlédnutí negativní formulace kmenu zadání; (d) uvedení pouze názvů organických látek / biomolekul (nikoliv vzorců) či (e) nestandardní zadání úlohy (úlohy netypické pro žáky středních škol).

5. Závěr

V článku bylo zmíněno sedm doporučení při sestavování testových uzavřených úloh s výběrem jedné správné odpovědi ze čtyř nabízených. Následně bylo uvedeno 12 příkladů úloh, jejichž zadání neodpovídala části doporučení pro tvorby testových úloh. Úlohy byly následně rozebrány včetně uvedení jejich základních charakteristik (úspěšnost, citlivost, analýza distraktorů).

Vzhledem k tomu, že úpravou zadání úloh je možné ovlivnit obtížnost a citlivost úloh tak, aby byly adekvátně obtížné a co nejvíce citlivé, autoři článku doporučují, aby pedagogové (tvůrci úloh) kladli nemalou pozornost při jejich sestavě s cílem vyhnout se v článku uvedeným konstrukčním či obsahovým chybám.

Článek byl sepsán s podporou projektu COOPERATIO v oblasti Subject Specific Education Research uděleného Univerzitou Karlovou.

LITERATURA

1. Chráska M.: *Didaktické testy*. Paido, Brno 1999.
2. Jeřábek O., Bílek M.: *Teorie a praxe tvorby didaktických testů*. Olomouc 2010.
3. George D., Mallery P.: *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4. vyd.). Allyn & Bacon, Boston 2003.
4. Gavora P.: *Úvod do pedagogického výzkumu*. Paido, Brno 2000.
5. Chvál M., Procházková I., Straková J.: *Hodnocení výsledků vzdělávání didaktickými testy*. Tiskárna bílý slon s.r.o., Plzeň 2015.
6. Teplá M., Šrámek M.: *Scientia in Educatione* 12(2), 32 (2022).
7. Varma S.: https://eddata.com/wp-content/uploads/2015/11/EDS_Point_Biserial.pdf, staženo 18. 1. 2023.
8. Štuka Č., Vejražka M.: *Testování a hodnocení studentů na VŠ*. Nakladatelství Karolinum, Praha 2021.

9. Šrámek M.: *Analýza přijímacích zkoušek z chemie na Přírodovědeckou Fakultu Univerzity Karlovy. Disertační práce*. Univerzita Karlova, Praha 2022.
10. Mareček A., Honza J.: *Chemie pro čtyřletá gymnázia – 1. díl*. Nakladatelství Olomouc s.r.o., Olomouc 2005.

M. Teplá and M. Šrámek (*Department of Teaching and Didactics of Chemistry, Faculty of Science, Charles University, Prague, Czech Republic*): **Tips and Recommendations for Creating Tests in Chemistry**

This article presents tips and recommendations for test creation (not only in chemistry). First, a general characteristics of tests, the individual items and distractors are defined. Subsequently, 7 tips and recommendations are given, shown on 12 test items. For each test item, the success rate, sensitivity and distractor attractivity are calculated. Finally, the impact of a low correlation coefficient between real and expected item difficulty is discussed.

Keywords: item analysis, sensitivity, difficulty, distractor attractivity, test quality

Acknowledgements

The support of the project COOPERATIO in the field of Subject Specific Education Research awarded by Charles University is gratefully acknowledged.