

Hvernig mæla á hugsmíðar með erlendum mælitækjum:

Pýðing og staðfærsla á spurningalistum og prófum

Sigurgrímur Skúlason
Námsmatsstofnun

Markmiðið með því að pýða og staðfæra mælitæki, til dæmis spurningalista eða próf, er að útbúa mælitæki sem aflar upplýsinga um hugsmíð með sama hætti í pýðingarlandi og frumútgáfa mælitækisins gerir. Ákveðnar próffræðilegar forsendur þarf að uppfylla til að það sé mögulegt. Í fyrri hluta greinarinnar er fjallað um próffræðilíkan sem dregur þessar forsendur fram en í síðari hluta er vinnuferli við pýðingu og staðfærslu lýst. Próffræðilíkanið er notað til að skilgreina nákvæmlega markmið með pýðingu og staðfærslu mælitækja og hvaða forsendur þurfa að vera til staðar svo að vel takist til. Líkanið hjálpar jafnframt við að draga fram afleiðingar af hnökrum eða ónákvæmni í pýðingu og skýrir um leið tilgang og mikilvægi ákveðinna verkhluta. Í síðari hluta greinarinnar verður ferlinu, þ.e. verkþáttum við pýðingu og staðfærslu á mælitækjum, lýst. Þessir verkþættir eru: Undirbúningur, pýðing og aðlögun texta, endurbætur, stöðlun (ef við á) og útgáfa. Byggt er á almennt viðurkenndu verklagi en hér er reynt að draga fram sjálfstæði verkþátta með skýrari hætti en áður hefur verið gert. Fjallað er stuttlega um verkþættina og þeir skilgreindir með það að markmiði að þeir sem þurfa að pýða og staðfæra mælitæki fái raunhæfar forsendur til að skipuleggja vinnuferlið.

Hagnýtt gildi: Pýðing og staðfærsla á mælitækjum, til dæmis spurningalistum eða prófum, lýtur öðrum lögmálum en aðrar pýðingar. Markmið slíkra pýðinga er að þýtt og staðfært mælitæki gefi sambærilegar upplýsingar og frumútgáfa þess en blæbrigði textans skipta ekki máli. Fjallað er um hvað einkennir próffræðilega eiginleika staðfærðra pýðinga og vinnuferlinu við pýðingu þeirra er lýst. Greinin er ætluð fagfólki og rannsakendum sem starfa á sviði félagsvísinda, menntamála, heilbrigðismála og fleiri sviðum og þurfa að nota staðfærðar pýðingar mælitækja eða standa sjálf að pýðingu og staðfærslu þeirra.

Hlutverk mælitækja er að afla upplýsinga. Í þessari grein verður fjallað um mælitæki þar sem tungumál í formi spurninga og svara er notað í mælingarferlinu. Slík tæki eru þróuð til að mæla tiltekna hugsmíð eða þekkingarsvið hjá ákveðnum markhópi (Crocker og Algina, 1986; Millman og Greene, 1989). Í mörgum fræðigreinum og starfi fagstétta eru slík mælitæki notuð til að safna upplýsingum með kerfisbundnum hætti. Þau eru notuð m.a. í félags- og hugvísindum, menntageiranum,

læknis- og hjúkrunarfræði og viðskiptafræði og mæla hugsmíðar eins og þunglyndi, sjálfsmynd, stjórnmalaskoðanir, neysluvenjur, heilbrigðissögu og félagslegan bakgrunn. Oft er talið æskilegt að flytja stöðluð greiningartæki, t.d. kunnáttu- og réttindapróf, spurningalista, mats- og gátlista og spurningaramma fyrir viðtöl milli málsvæða, en hlutur tungumáls í slíkum mælitækjum gerir það erfitt.

Pýðingu og staðfærslu verður að haga með þeim hætti að upplýsingar sem fást

með staðfærðri þýðingu mælitækisins hafi sambærilega próffræði- og tölfræðilega eiginleika og frumútgáfa þess (Hulin, 1987; Hambleton og Bollwark, 1991; International Test Commission, 2001; van de Vijver og Hambleton, 1996). Þetta markmið gerir þýðingar mælitækja frábrugðnar öðrum þýðingum. Ekki er ætlunin að ná fram ákveðinni upplifun lesanda eða áferð texta eins og þegar bókmenntatexti er þýddur. Það er heldur ekki markmið að upplýsingar komist orðrétt til skila eins og þegar tæknilegur leiðbeiningatexti er þýddur (Heimir Pálsson og Höskuldur Þráinsson, 1988). Í staðfærðri þýðingu mælitækis er lögð áhersla á að afla sambærilegra upplýsinga, að niðurstöður hafi sömu próffræðilegu og tölfræðilegu eiginleika og í frumútgáfu (International Test Commission, 2001).

Próffræðilegir eiginleikar mælitækisins byggjast á eiginleikum einstakra atriða, t.d. þyngd, aðgreiningargetu og tengslum við önnur atriði (Feldt og Brennan, 1989; Gulliksen, 1950/1987). Orðið *atriði* er hér notað um spurningar eða staðhæfingar, leiðbeiningar sem þeim fylgja og leiðir til að svara spurningum. Hverju atriði er ætlað að snerta á ákveðnum fleti hugsmíðarinnar og saman afla þau upplýsinga um hugsmíðina sem mæld er (Crocker og Algina, 1986; Millman og Greene, 1989). Þýðing og staðfærsla felur í sér að texti mælitækisins er þýddur og aðlagður menningu með þeim hætti að próffræði- og tölfræðilegir eiginleikar halda sér. Hugsmíðar geta verið breytilegar á milli mál- eða menningarsvæða þv ví tengslin milli efniviðar atriðis, þess sem spurt er um eða staðhæft í atriðinu, og hugsmíðarinnar geta reynst ólík frá einu menningarsvæði til annars. Við það geta komið upp vandamál í þýðingu og staðfærslu. Sem dæmi má nefna mælingar á magnbundinni hugsun barna við upphaf grunnskóla. Í bandarískri útgáfu mælitækis hefur atriði er lýtur að innbyrðis verðgildi peninga mjög sterk tengsl við þessa hugsmíð en tengsl eru engin í íslenskri þýðingu mælitækisins (Sigurgrímur Skúlason, 2004a, 2004b). Í tilvikum sem þessu

þarf að finna annan efnivið í atriðið sem gerir kleift að mæla ákveðinn flöt hugsmíðarinnar.

Algengt er að mælitæki séu þýdd og staðfærð til að nota við rannsóknir eða greiningar hér á landi en líklega vantar nokkuð upp á þekkingu á aðferðum við þýðingu og staðfærslu. Þessari grein er ætlað að bæta þar úr með tvennum hætti. Í fyrri hluta greinarinnar verður kynnt próffræðilíkan sem gerir kleift að skilgreina í hverju markmið þýðingar og staðfærslu mælitækja felast. Í síðari hluta hennar verður vinnuferli við þýðingu og staðfærslu lýst með einföldum hætti sem fjórum eða fimm aðskildum verkþáttum. Byggt er á viðurkenndum vinnuferlum en útfærsla heildarferlis er hér nokkuð frábrugðin því sem almennt tíðkast (Brachen og Barona, 1991; Brislin, 1970, 1986; Hambleton, 1993; Hambleton og Bollwark, 1991; International Test Commission, 2001; Prieto, 1992) og er ætlunin að gera hlutverk hvers verkþáttar sem ljósast.

Próffræði staðfærðrar þýðingar

Eins og áður kom fram eru mælitæki þróuð til að safna upplýsingum um skilgreinda hugsmíð og við þróun mælitækisins eru byggðir inn í það ákveðnir próffræðilegir eiginleikar sem ráða því hvernig mælitækið höndlar hugsmíðina (Crocker og Algina, 1986). Svo að staðfærð þýðing mæli hugsmíð með sama hætti og frumútgáfa verða þessir eiginleikar að haldast gegnum þýðingarferlið (International Test Commission, 2001).

Vandamál í þýðingu mælitækja geta falist í kerfisbundinni skekkju tengdri einstökum atriðum (*item bias*) eða hugsmíðum (*construct bias*) (Ellis, 1989; Hambleton, 1993; Hambleton og Bollwark, 1991; Hulin, 1987). Eitt nákvæmasta líkanið af þessu tagi er byggt á svarferlalkani (*item response model*) (Hulin, 1987), og byggir á tengslum milli færni eða stöðu einstaklings á hugsmíðinni sem verið er að mæla og líkum á að hann svari einstökum atriðum með ákveðnum hætti. T.d. getur einstaklingur sem hefur

jákvæða afstöðu til náttúruverndar verið mjög ósammála fullyrðingu um að framkvæmd sem veldur umhverfisraski væri æskileg. Eru tengslin skilgreind út frá einum til þremur próffræðilegum eiginleikum atriðanna, auk stöðu þátttakenda á heildarkvarða. Hulin (1987) telur þýtt atriði sambærilegt frumútgáfu þegar þessir eiginleikar þess eru sambærilegir frumútgáfu. Í öllum þessum líkönum felst að einangra má hnökra í þýðingu og staðfærslu mælitækja með því að finna atriði sem hafa ólíka próffræðilega eiginleika í þýðingu og frumútgáfu, t.d. þyngd eða aðgreiningarhæfni.

Poortinga og van de Vijver (van de Vijver og Poortinga, 1997, 2005) hafa sett fram víðtækari kenningu um samræmi milli staðfærðrar þýðingar og frumútgáfu. Þeir gera ráð fyrir fjórum stigum samræmis: Að einstök atriði séu sambærileg í þýðingu og frummáli, að heildarkvarði eða undirkvarðar séu sambærilegir, að hugsmíð sé mæld með sama hætti í báðum útgáfum og að mælieiningar séu sambærilegar. Ákveðin samsvörun er milli próffræðilíkansins sem hér verður sett fram og þessarar kenningar um sambærilegar þýðingar.

Líkanið sem lýst er í þessari grein er byggt á grunni klassíska próffræðilíkansins (*classical test theory*) (Gulliksen, 1950/1987) og viðunandi frávik eða hnökrar í staðfærðri þýðingu skilgreindir út frá heildarstigatölu, stigatölum undirþátta eða út frá einstökum atriðum. Heildarstigatölu mælitækis má lýsa með grunnjöfnu klassíska próffræðilíkansins. Samkvæmt henni eru mæligildi eða heildarstig, x , (*test score*) samsett af markgildi, t , (*true score*) og óvissuþætti, e_t , (*error*),

$$x = t + e_t.$$

Mæligildi eða heildarstig einstaklings er niðurstaðan sem mælitækið gefur. Markgildi má líta á sem besta mat á hugsmíðinni sem þetta mælitæki getur gefið. Einnig má líta á það sem

niðurstöðu sem fengist ef unnt væri að mæla hugsmíðina án óvissuáhrifa. Óvissuþáttur felst í tilviljanakenndum áhrifum sem tengjast ekki hugsmíðinni sem verið er að mæla (Feldt og Brennan, 1989; Gulliksen, 1950/1987). Með sama hætti má skoða dreifingu á mælitækinu, σ^2 ,

$$\sigma_x^2 = \sigma_t^2 + \sigma_{e_t}^2.$$

Hér táknar σ_x^2 dreifingu heildarstiga. Hún er summa liðanna σ_t^2 , sem er dreifing markgildis og $\sigma_{e_t}^2$, sem er dreifing óvissuþátta.

Þegar mælitæki er þýtt og staðfært er markmiðið að endurskapa mæligildið og dreifingu á því í þýðingarlandinu. Um leið opnast leiðir fyrir dreifingu tengda óvissu-áhrifum¹ vegna hnökra í þýðingu og menningarlegra þátta. Þau eru táknud með σ_{ep}^2 og stendur p fyrir þýðingu og staðfærslu. Dreifing á staðfærðri þýðingu mælitækis reynist því

$$\sigma_x^2 = \sigma_t^2 + \sigma_{e_t}^2 + \sigma_{ep}^2.$$

Eftir því sem óvissuþátturinn sem tengist þýðingu og staðfærslu, σ_{ep}^2 , verður stærri dregur úr samræmi milli staðfærðrar þýðingar og frumútgáfu. Það fer að miklu leyti eftir því hvernig tekst að halda þessum óvissuþætti í lágmarki hvort staðfærð þýðing verður viðunandi. Til að þetta markmið náist getur reynst nauðsynlegt að laga efnistöð atriða að menningarlegum aðstæðum, m.ö.o. breyta því sem spurt er um, til að þau gefi sambærilegar upplýsingar um hugsmíðina og frumútgáfan (International Test Commission, 2001; van de Vijver og Poortinga, 1997, 2002).

Það er algengt að mælitæki hafi nokkra undirþætti og þá er hverjum þeirra ætlað að meta ákveðinn flöt hugsmíðarinnar. Lýsa má dreifingu á mælitæki með k skilgreinda undirþætti sem

$$\sigma_x^2 = \sigma^2_1 + \sigma^2_2 + \dots + \sigma^2_k + \sigma_{e1}^2 + \sigma_{e2}^2 + \dots + \sigma_{ek}^2 + \sigma_{12} + \sigma_{13} + \dots + \sigma_{(k-1)k},$$

þar sem liðir auðkenndir með tölum tákna dreifingu undirþátta, liðir auðkenndir með e og

¹ Í klassíska próffræðilíkaninu er villuþáttur ekki greindur í liði. Hér er vikið frá þessu til að skýra villuáhrifin. Tæknilega séð er hér farið yfir á svið alhæfingakenninga (*generalizability theory*) þar sem unnt er að brjóta villuþátt niður eftir uppruna eða áhrifaþáttum (Feldt og Brennan, 1989).

tölu tákna óvissuþátt þeirra og liðir auðkenndir með tveimur tölum (t.d. σ_{12}) sýna samdreifingu (*covariance*) undirþátta, en samdreifing geymir upplýsingar um tengsl milli undirþátta og er skyld fylgni. Við þýðingu og staðfærslu geta þýðingarbundnir óvissuþættir tengst hverjum undirþætti, σ_{uv}^{2ep} . Til að staðfærð þýðing hafi sömu eiginleika og frumútgáfa verður dreifing á undirþáttum, σ_u [Frumútgáfa] \approx σ_u [Staðfærsla], að vera með sama hætti í báðum útgáfum og einnig samdreifing undirþátta, σ_{uv} [Frumútgáfa] \approx σ_{uv} [Staðfærsla]², og óvissuþættir tengdir þýðingu og staðfærslu verða að vera í lágmarki. Þar sem fá atriði eru í hverjum undirþætti er nauðsynlegt að haga þýðingu og staðfærslu atriða þannig að dreifing hvers atriðis skarist við ólíkar hliðar hugsmíðarinnar með sama hætti og frumgerð þess. Ef til dæmis þýðing atriðis sem metur þá hlið hugsmíðarinnar sem samsvarar undirþætti u er þannig að dreifing á þessum undirþætti verði frábrugðin því sem gerist í frumútgáfu verður samdreifing þar sem undirþáttur u á í hlut ávallt frábrugðin því sem gerist í frumútgáfu. Hætt er við að þetta hafi þau áhrif að þáttabygging mælitækisins riðlist. Algengasta viðmið um hve vel þýðing og staðfærsla heppnast er að þáttabygging þýðingar sé hin sama og frumútgáfu (Gierl, 2000; van de Vijver og Poortinga, 2002), tölfræðileg forsenda þess að þáttabygging verði sambærileg er að samdreifing sé sambærileg (Bollen, 1989; McDonald, 1985).

Að síðustu má bera saman próffræðilega eiginleika einstakra atriða í staðfærðri þýðingu og frumútgáfu. Jafna fyrir slíkt líkan yrði sambærileg þeirri sem lýsir undirprófum. Í stað liða fyrir undirpróf kæmi þá einn liður fyrir dreifingu á hverju atriði auk liða sem sýna samdreifingu atriðanna. Eins og áður verður dreifing á einstökum atriðum að vera um það bil jöfn (σ_{i}^2 [Frumútgáfa] \approx σ_{i}^2 [Staðfærsla])³ og samdreifing þeirra einnig (σ_{ij} [Frumútgáfa] \approx σ_{ij} [Staðfærsla]) til þess að staðfærð þýðing mæli hugsmíðina með sama hætti og frumútgáfa. Eins og áður geta óvissuþættir vegna þýðingar

og staðfærslu tengst hverju atriði. Þetta sýnir að við þýðingu og staðfærslu er nauðsynlegt að vanda til hvers atriðis því það eru í raun próffræðilegir eiginleikar atriðanna sem ráða því hverjir próffræðilegir eiginleikar staðfærðrar þýðingar sem heild verða (Felt og Brennan, 1989). Þetta þýðir að við þýðingu og staðfærslu er nauðsynlegt að velja efni hvers atriðis þannig að tengsl þess við hugsmíð séu svipuð og í frumútgáfu. Ef þýðing og staðfærsla atriðis hleypir menningarlegum eða þýðingarbundnum óvissuþáttum inn í dreifingu atriðisins hefur það einnig áhrif á samdreifingu þegar atriðið á hlut að. Slík þýðingaráhrif eru því ekki aðeins bundin við atriðið sjálft heldur smitast þau inn í aðra hluta mælitækisins.

Mikilvægi þess sem fram hefur komið að ofan má skýra með dæmum. Ef efnistöð atriðis eiga ekki við í þýðingarlandi getur það gerst að nær allir svari með sama hætti (rétt eða rangt, oftast eða aldrei, mjög sammála eða mjög ósammála. Í áður nefndri rannsókn á skilningi sex ára barna svöruðu nær allir leik- og grunnskólakennarar að börnin hefðu ekki vald á verðgildi peninga og þetta atriði legði því nær ekkert til dreifingar á undirþætti sem mældi hugsmíðina magnbundna hugsun. Jafnframt hvarf samdreifing sem í frumútgáfu tengdi þennan þátt við hugsmíðina almenna þekkingu. Við þetta urðu báðir undirþættir óskýrir í þáttagreiningu (Sigurgrímur Skúlason, 2004a). Hér þarf að finna annað atriði sem tengist magnbundinni hugsun í daglegu lífi og nota í stað verðgildis peninga til að staðfærð þýðing mæli sömu hugsmíð og frumútgáfan.

Nefna má annað dæmi um óviðeigandi efni við atriðis. Ef atriði í frumútgáfu spurningalista sem ætlað er að mæla öryggi nemenda í skólum lyti að því hvort vopnaburður, t.d. að koma með byssu í skólann, sé algengur meðal nemenda myndu flestir íslenskir nemendur svara því til að það gerðist aldrei. Sumir þeirra myndu hinsvegar svara út í hött undir áhrifum frá kvikmyndum eða sjónvarpsefni. Dreifing á atriðinu réðist þá eingöngu af óvissuáhrifum

² Bókstafirnir u og v tákna hér „undirpróf“ og „eitthvert annað undirpróf“.

³ Bókstafirnir i og j tákna hér „prófatriði“ og „eitthvert annað prófatriði“.

tengdum þýðingu og staðfærslu, σ_{ep}^2 . Við þetta dregur atriðið dreifingu sem ekki tengist hugsmíð (*construct irrelevant variance*) inn í mælitækið. Óbreytt nær atriðið ekki að meta þann flöt hugsmíðarinnar sem því var ætlað en veldur á sama tíma skaða með þeirri dreifingu sem það dregur inn (Messick, 1989). Að auki er líklegt að atriðið dragi úr trausti íslenskra unglinga á rannsókninni. Til þess að meta öfgakenndustu hliðar hugsmíðarinnar þyrfti að sníða efnistöð atriðisins að veruleika markhóps. Ávallt verður þó að meta þörfina á að breyta efnistöðum einstakra atriða í ljósi væntanlegrar notkunar mælitækisins, t.d. við greiningu, rannsóknir, mat á aðstæðum í þýðingarlandi eða alþjóðlegan samanburð. Oftast gefa niðurstöður forprófana bestu upplýsingar um hvaða leið er heppilegust við að mæla hugsmíðina.

Vinnuferli við þýðingu og staðfærslu á mælitæki

Þýðingu og staðfærslu verður hér lýst sem vinnuferli er felur í sér fjóra eða fimm verkþætti: Undirbúning, þýðingu texta, endurbætur og gæðamat, stöðlun og útgáfu. Margvísleg handtök er lúta að undirbúningi eru hér tekin saman í einum verkþætti. Sjálf þýðing og staðfærsla textans er unnin í næsta verkhluta. Þriðji verkhluti felst í endurbótum þar sem staðfærð þýðing er slípuð til. Stöðlun felst í að útbúa viðmið sem gera kleift að túlka niðurstöður með tilvísun í vel skilgreindan markhóp. Þessi verkhluti á einkum við um stöðluð greiningar- og matstæki. Lokahluti, hér kallaður útgáfa, felur í sér að ganga frá mælitækinu í endanlegri gerð, rannsaka eiginleika þess og kynna það notendum, þeim sem nota mælitækið. Verkhlutunum verður nú lýst. Byggt er á eldri umfjöllunum (Brachen og Barona, 1991; Brislin, 1970, 1986; Hambleton, 1993, 2005; Hambleton og Bollwark, 1991; International Test Commission, 2001; Prieto, 1992) en skipting verkferlisins og framsetning er með nýstárlegu sniði.

Undirbúningur

Áður en ákvörðun um þýðingu og staðfærslu er tekin hefur óformlegur undirbúningur átt sér stað. Hætt er við að sú vinna hafi ekki verið nógu markviss og því er nauðsynlegt að endurmeta þörf og forsendur fyrir þýðingu og staðfærslu mælitækis þegar ákvörðun um verkefnið er í vinnslu (International Test Commission, 2001). Til dæmis verður að kanna hver raunveruleg þörf fyrir mælitæki er og hvernig mælitæki svarar best þeim þörfum. Meta þarf styrkleika og veikleika mælitækja sem standa til boða. Sömu áherslur eiga hér við og þegar valið er mælitæki til greiningar eða í rannsókn (Crocker og Algina, 1986). Mælitækið verður að mæla hugsmíðina í samræmi við skilgreiningu hennar í þýðingarlandinu. Því verður að athuga hvort eitthvað standi út af eða mælitækið feli eitthvað í sér sem ekki á við. Hlutverk mælitækisins verður að henta tilgangi (rannsókn, skimun, greiningu), próffræðilegir eiginleikar verða að henta notkun og slíkir eiginleikar frumútgáfu (t.d. réttmæti, áreiðanleiki eða greiningarhæfni) þurfa að vera fullnægjandi (Brachen og Barona, 1991; Hambleton og Bollwark, 1991; International Test Commission, 2001; Prieto, 1992).

Full ástæða er til að huga að því hvort betur fari á því að þróa nýtt mælitæki en þýða og staðfæra erlent. Meðal þess sem ræður því er hve sterk menningarleg sérkenni hugsmíðarinnar eru, hvort völ sé á góðu erlendu mælitæki, hvort hugsmíðin sé nátengd menningarlegum sérkennum og hversu flókið mælitækið er (van de Vijver og Poortinga, 2005). Þótt til séu erlend mælitæki geta þau verið ófullnægjandi. Til dæmis er algengt að mælitæki sem komin eru til ára sinna séu ekki í samræmi við fræðilega þróun eða skilgreiningar á hugsmíðum. Af þessari ástæðu getur verið æskilegt að þróa nýtt mælitæki í stað þess að þýða og staðfæra erlent (International Test Commission, 2001; van de Vijver og Poortinga, 2005).

Þegar búið er að velja mælitæki til þýðingar og staðfærslu er nauðsynlegt að huga að öllum þeim verkþáttum sem fjallað er um í framhaldinu og skipuleggja vinnuferlið eftir því

sem kostur er (International Test Commission, 2001).

Þýðing og staðfærsla á texta mælitækis

Þessi verkhluti lýtur að vali á þýðendum og vinnu við þýðingu og staðfærslu á texta mælitækisins. Þýðendur eru að sjálfsögðu valdir eftir færni og þekkingu sem æskileg er og vinnulag mótast af því hvort ein eða tvær þýðingar eru gerðar og hvernig gengið er frá lokaútgáfu textans þegar tveir þýða (Grisay, 2003, 2004; van de Vijver og Hambleton, 1996; Maxwell, 1996).

Viðfangsefni þýðenda er að þýða texta, atriði og leiðbeiningar, sjá til þess að málfar sé við hæfi markhópsins og inntak atriða sé í samræmi við aðstæður hans. Vinnulag getur ýmist verið með þeim hætti að einn þýðandi vinni verkið eða tveir þýðendur geri hvor sína þýðingu. Þegar tvær sjálfstæðar þýðingar eru gerðar sér þriðji þýðandi eða vinnuhópur um að ganga frá lokatexta á grunni beggja þýðinganna (Grisay, 2003, 2004; Maxwell, 1996). Kostir þess að tveir þýði eru að þeir geta dregið fram ólíkar hliðar á viðfangsefninu, farið ólíkar leiðir í orðalagi eða efnistökom þegar atriði er breytt. Þessi atriði geta verið málfarsleg, t.d. val á orðum, þýðing margræðra orða, markviss framsetning á góðri íslensku og fleira. Jafnframt koma fram ólíkar leiðir til að færa efni atriða að íslenskum aðstæðum og ólíkar leiðir til að finna nýjan efnivið þegar efniviður í frumútgáfu á ekki við á Íslandi. Þegar lokaútgáfa er gerð upp úr báðum þýðingum blasa við atriði sem á milli ber og þriðji þýðandi eða vinnuhópurinn velur þá útfærslu sem betri þykir eða kemur með nýja tillögu. Þegar ákveðið er hvort einn eða tveir þýðendur komi að verkinu er nauðsynlegt að huga að því hvaða færni er mikilvægust fyrir verkið og hvort vól sé á þýðanda sem hefur allt til að bera. Einnig verður að meta hvort ásættanlegt sé að fórna þeim ávinningi sem fæst með því að tveir þýði (Grisay, 2003, 2004; Maxwell, 1996).

Val þýðenda verður að byggja á því hver markmið þýðingarinnar eru. Æskilegt er að

þýðendur hafi færni á fimm sviðum, þ.e. þýðandi verður að hafa gott vald á báðum tungumálum, hafa góða þekkingu á menningarlegum sérkennum beggja samfélaga, á hugsmíðinni sem mæld er, á markhópnum og þekkingu á prófagerð. Æskilegt er að velja þýðendur sem hafa sem flesta þessara þátta á valdi sínu og forgangsraða ólíkri færni eftir mikilvægi hverju sinni (Hambleton og Bollwark, 1991; van de Vijver og Hambleton, 1996).

Þessi atriði vega misþungt eftir aðstæðum hverju sinni. Færni í báðum tungumálum er undantekningarlaust nauðsynleg til að skynja blæbrigði í frumútgáfu og skila þeim á lipru máli í þýðingu (Grisay, 2004; Heimir Pálsson og Höskuldur Þráinsson, 1988). Mýmörg dæmi um stirt orðalag sem gerir atriði flókin eða illskiljanleg, t.d. ensk setningaskipan, benda til að góðri færni í íslensku sé ekki síður ábótavant hjá þeim sem þýtt hafa mælitæki hér á landi en færni í frummáli.

Þýðandi verður bæði að þekkja til hugsmíðar sem mæld er og til markhópsins, þeirra sem ætlað er að svara mælitækinu. Þetta er mikilvægt til að meta hvenær efnistöku atriðis (hegðun, staðhæfing, skoðun) henta ekki til að mæla sama flöt hugsmíðar í báðum löndum og hvenær þarf að velja annan efnivið. Þessi þekking er þýðanda einnig nauðsynleg til að velja orðafar í takt við tungutak og þankagang markhóps, til dæmis að nota einfalt og hlutbundið orðalag þegar við á, nota faglegt orðfæri einungis fyrir sérmenntaðar fagstéttir, meta hvenær nota má slangur sem beitt er hjá markhópi og nota hluti eða atburði úr hversdagslegu umhverfi þegar efnivið atriðis er breytt (Grisay, 2004; Hambleton, 1993, 2005; Hulin, 1987; International Test Commission, 2001). Þekking á hugsmíð er einnig nauðsynleg til að beiting faglegs orðafars verði markviss (van de Vijver og Hambleton, 1996). Líta má svo á að þetta falli undir yfirborðsréttmæti (*face validity*) og getur verið nauðsynleg forsenda þess að þeir sem svara taki mælitækið alvarlega og svari af heiðarleika (Messick, 1989).

Æskilegt er að þýðendur séu vel að sér um gerð prófa og spurningalista og um orðalag

spurninga eða prófatriða. Þeir verða að tryggja að málfar, spurning, fyrirmæli og form falli saman, en einnig er mikilvægt að halda formi prófatriða óbreyttu. Leiðbeiningar um markviss vinnubrögð við prófagerð og samningu prófatriða eða spurninga eiga almennt við um þýðingar og staðfærslur á mælitækjum (sjá m.a. Crocker og Algina, 1986; Fowler, 2000; Grisay, 2004; Þorlákur Karlson, 2003).

Endurbætur og gæðamat

Rannsóknir á þýðingum og staðfærslum mælitækja hafa sýnt að jafnvel þegar vel er að verki staðið er full þörf á að leita hnökra í staðfærðri þýðingu og bæta úr ef þeir finnast. Hafa þær leitt í ljós að margvísleg vandamál geta komið upp. Efniviður eða hugmynd atriðisins getur brenglast í þýðingu, framsetning þess verið í ósamræmi við það sem markhópur á að venjast og þýðing og staðfærsla getur dregið úr eða bætt við merkingu. Algennt orð er stundum þýtt með sjaldgæfu orði eða öfugt, þýðing verður lengri en frumtexti, málfar fylgir stundum reglum frummáls, setningarbygging tekur stundum ekki mið af málhefð þýðingarmáls eða þroska væntanlegs markhóps, ósamræmi er í þýðingu orða sem koma oft fyrir í mælitækinu, efniviður atriðis á ekki við í þýðingarlandi, vinnureglur um uppbyggingu atriða í prófum eða spurningalistum eru brotnar og uppsetning er ekki í samræmi við frumútgáfu (Allouf, Hambleton og Sireci, 1999; Ercikan, 1998; Gierl og Khaliq, 2001; Solano-Flores, Contreras-Niño og Backhoff, 2005).

Hnökra á staðfærðri þýðingu mælitækisins veikja hana og leiða til þess að próffræðilegir eiginleikar verða ekki sem skyldi eins og sýnt var fram á fyrir í þessari grein. Til að geta metið hvort nauðsynlegt sé að breyta atriðum í staðfærðri þýðingu mælitækisins er nauðsynlegt að safna upplýsingum um styrkleika og veikleika hennar (Brachen og Barona, 1992; Hambleton og Bollwark, 1991; Prieto, 1992). Huga þarf að því hvort atriði nái að meta sama flöt hugsmíðar og frumútgáfan gerir, hvort efnistöð atriða henti markhópi,

hvort málfar sé viðeigandi fyrir markhóp, hvort textinn sé á góðri íslensku, hvort próffræðilegir eiginleikar atriða séu heppilegir og hvort tengsl mælitækisins við aðrar breytur séu í samræmi við væntingar (Hambleton, 2005; van de Vijver og Hambleton, 1996). Mismunandi aðferðir við gæðamat henta til að kanna þessi atriði og því er nauðsynlegt að nota upplýsingar úr ólíkum áttum til að draga fram eins marga vankanta og unnt er. Því verður að velja markvissar leiðir við gæðamat og verja fjármunum og tíma með sem bestum hætti.

Til að fá upplýsingar um hve vel efniviður atriða hentar markhópi er best að leita til hópsins sjálfs. Einfaldast er að láta fáeina einstaklinga svara mælitækinu og benda á atriði sem þeim þykja illa orðuð eða torskilín (Brachen og Barona, 1992; Hambleton og Bollwark, 1991; Prieto, 1992). Gagnleg aðferð sem gerir kleift að fá fram fjölbreytilegar athugasemdir markhóps er rýnihópar (*focus groups*) (Krueger og Casey, 2000). Spurt er út í einstök atriði í fjögurra til átta manna hópum og umræðum komið af stað þar sem ólíkum skilningi eða sjónarmiðum er velt upp. Hægt er að beina umræðunni að einstökum atriðum mælitækisins eða að hugsmíðinni eins og þátttakendur skynja hana og ræða atriðin í ljósi hennar.

Ábendingar um málfar er best að fá frá markhópi og margar leiðir til þess færar. T.d. er heppilegt fyrsta skref í gæðamati að láta einstaklinga úr markhópi svara mælitækinu og benda á atriði sem stinga í stúf. Þessi atriði eru síðan lagfærð áður en farið er í umfangsmeira gæðamat (Brachen og Barona, 1992; Hambleton og Bollwark, 1991; Prieto, 1992). Önnur aðferð er að láta þátttakendur „hugsa upphátt“ meðan þeir svara mælitækinu. Þegar þeir orða hugrenningar sínar kemur fram hvað þeir hnjóta um eða skilja illa (Grisay, 2003). Í þriðja lagi má nýta rýnihópa með þátttakendum úr markhópi, séu þeir notaðir, og beina hluta umræðunnar að málfari og því hvernig þátttakendur skilja texta eða atriði mælitækisins. Í fjórða lagi má fá fjölbreytilegar athugasemdir frá þátttakendum

í forprófunum þegar haldið er vel utan um spurningar, athugasemdir og viðbrögð þeirra. Styrkur þessa er að fleiri þátttakendur koma að forprófunum en í öðrum aðferðum (Grisay, 2004; Hambleton, 2005). Þeir sem þekkja vel til markhópsins geta einnig bent á hvernig best er að haga orðalagi og málfari þannig að það sé í samræmi við það sem gerist hjá markhópi eða á fagsviðinu. Þeir bera einnig saman frumútgáfu og staðfærða þýðingu. Mikilvægt er að fá sérfræðinga sem ekki hafa áður komið að þýðingu og staðfærslu mælitækisins því hætt er við að þeim sem þegar hafa unnið að því sjáist yfir veikleika hennar (Brachen og Barona, 1992; Hambleton og Bollwark, 1991; Prieto, 1992).

Allajafna gefa sérfræðingar og fagfólk sem starfar í tengslum við markhóp gagnlegastar upplýsingar um hvort þýddar og staðfærðar útgáfur atriða tengjast hugsmíð með sama hætti og frumútgáfan gerir (Brachen og Barona, 1992; Hambleton, 2005; Hambleton og Bollwark, 1991; Prieto, 1992). Þessir aðilar fá það hlutverk að meta hvert atriði með tilliti til þess hve vel það hentar, hvaða flöt hugsmíðar það muni tengjast. Gagnlegt er fyrir þessa aðila að hafa upplýsingar um hlutverk atriðanna, t.d. hvaða undirkefna atriðið tengist. Mikilvægt er að fá ábendingar um atriði sem sérfræðingum og fagfólki sýnist að muni lítt tengjast hugsmíð. Þó fram komi ábendingar af þessu tagi er ekki ljóst hvort menningarbundinn munur er á tengslum atriðis og hugsmíðar eða hvort óþekktar ástæður liggja að baki slíkum tengslum (Hambleton og Bollwark, 1991; Prieto, 1992). Því er æskilegt að forprófa þýðingu atriðisins auk annarrar hugsanlegrar staðfærslu á því og velja á milli ólíkra leiða við að staðfæra atriðið eftir niðurstöðum forprófana. Markhópur getur í mörgum tilvikum einnig gefið mikilvægar upplýsingar um tengsl atriða við hugsmíð. Þegar rýnihópar eru notaðir má til að mynda spyrja um hvaða merkingu ákveðið orðalag hafi eða spyrja um skoðun, hegðun eða atburð sem hugsanlegt væri að nota sem efnivið í atriði.

Mikilvægt er að fá gagnrýni íslenskufræðinga á málfar staðfærðrar þýðingar. Algengt er að aðrir sérfræðingar bendi á málvillur en hlutverk íslenskufræðings er að benda á málfarsatriði, setningaskipan eða stúlbrigði sem færa má til betri vegar, þó ekki sé beinlínis rangt með farið, til að textinn renni vel á lipurri íslensku.

Upplýsingar um próffræðilega eiginleika atriða, t.d. þyngd, aðgreiningarhæfni, dreifitölur og samdreifingarliði eða tengsl við önnur atriði og ytri breytur, fást einungis með forprófunum. Mikilvægt er að úrtak sem stuðst er við í forprófunum komi úr væntanlegum markhópi (Crocker og Algina, 1986). Stærð úrtaks skiptir miklu um niðurstöður forprófana og er ávallt betra að þátttakendur séu fleiri en færri. Þyngd er einn stöðugasti eiginleiki atriða og fáeinir tugir þátttakenda gefa upplýsingar um þyngd atriða. Þegar úrtak nær u.þ.b. 100 þátttakendum er unnt að huga að niðurstöðum um tengsl atriða við mælitækið, innbyrðis þyngdarröð eða tengslum við aðrar breytur. Nokkur hundruð þátttakendur þarf hins vegar til að kanna þáttabyggingu eða skoða þyngd eða fylgni eftir undirhópum, og um 500 til 800 þarf til að kanna einkenni svarferlálíkana. Almennt yfirlit um algengustu aðferðir við atriðagreiningu má m.a. finna hjá Crocker og Algina (1986).

Þyngd prófatriða eða hlutfall svara á svarmöguleika gefur mikilvægar upplýsingar um próffræðilega eiginleika staðfærðrar þýðingar í heild (Crocker og Algina, 1986). Gagnlegt er að nota aðferðir tengdar atriðabundnum hópamun (*differential item functioning*) til að bera þyngd þýddra atriða saman við þyngd í frumútgáfu og einnig athuga hvort meðaltöl skilgreindra hópa fylgi sama mynstri og í frumútgáfu (t.d. eftir aldri eða kyni, van de Vijver og Hambleton, 1996). Atriðabundinn hópamunur felst í að atriði hafi ólíka próffræðilega eiginleika í skilgreindum hópum og að breytileikinn sé óháður getumun (Thissen, Steinberg og Gerrard, 1986). Algengustu aðferðir á þessu sviði, Mantel-Haenszel kí-kvadrat og SIBTEST, henta illa í þýðingar- og staðfærsluverkefnum því úrtök í forprófunum

eru almennt minni en til þarf og nauðsynlegt er að hafa upplýsingar um svör þátttakenda við atriðum í báðum útgáfum (Clauser og Mazor, 1998). Aðferðir sem ráða við lítil úrtök og styðjast einungis við þyngd atriða eru raunhæfari valkostir, til dæmis Delta-rit (*delta-plot*) (Angoff, 1982) og lagskipt *p*-gildi (*conditional p-value*, Dorans, 1989). Nýlegt yfirlit um þessar aðferðir má finna í Sireci, Patsula og Hambleton (2005).

Þegar unnið er með Likert-kvarða er mikilvægt að huga að því hvernig svör dreifast á kvarðana (Crocker og Algina, 1986). Í sumum tilvikum beita samsvarandi markhópar svarkvörðum með ólíkum hætti eftir menningarsvæðum. Dæmi eru um að menningarmunur felist í hvernig jákvæðar staðhæfingar um þátttakanda eru túlkaðar (Baumgartner og Steenkamp, 2001; Tanzer, 1995), hvernig sterkustu afstöðupunktur eru notaðir (Arce-Ferrer, 2005; van Herk, Poortinga og Verhallen, 2004), eða að staðhæfingar með jákvæðri og neikvæðri afstöðu séu túlkaðar með ólíkum hætti (Baumgartner og Steenkamp, 2001; Wong, Rindfleisch og Burroughs, 2003). Vegna þessa getur verið erfitt að flytja mælitæki á sumum hugsmíðum milli menningarsvæða.

Mikilvægt er að skoða innbyrðis tengsl atriða eða undirprófa þegar á forprófunarstigi til að geta brugðist við ef staðfærð þýðing mælir hugsmíð með öðrum hætti en frumgerð mælitækisins gerir (Gierl, 2000; van de Vijver og Leung, 1997). Þáttgreining er aðferð sem lítur á innbyrðis tengsl allra atriða samtímis, helstu afbrigði hennar eru hefðbundin þáttgreining (*factor analysis*) og stýrð þáttgreining (*confirmatory factor analysis*, McDonald, 1985). Síðari leiðin er betri kostur því hún gerir kleift að skilgreina líkan af byggingu mælitækisins, út frá fræðilegu líkani eða tengslum í frumútgáfu mælitækis, og prófa hvort innbyrðis tengsl atriða reynist í samræmi við það líkan (Bollen, 1989; van de Vijver og Poortinga, 2002). Þegar úrtök í forprófunum eru lítil reynist ekki unnt að kanna byggingu mælitækisins með markvissum hætti fyrir en úrtökum vegna stöðlunar eða notkunar er

safnað. Þá er æskilegt að skoða innbyrðis fylgni einstakra atriða eða undirprófa í leit að tengslum sem reynast óeðlilega veik eða sterk.

Gagnlegt er að kanna tengsl þýðingar og staðfærslu við breytur sem fræðilegur bakgrunnur hugsmíðar bendir til að mælitækið tengist, t.d. kyn, menntun, aldur eða breytur sem hugsmíð tengist. Upplýsingar um slík tengsl eru mikilvægar til að meta gæði staðfærðrar þýðingar og réttmæti hennar (van de Vijver og Leung, 1997; van de Vijver og Poortinga, 1997). Sjaldan er unnt að afla upplýsinga um tengsl við slíkar breytur að nokkru marki í forprófunum, en sé það unnt getur mikið áunnist.

Stöðlun

Þegar unnið er með stöðluð greiningartæki er stöðlunarhluti einn þáttur ferlisins. Gögnum til að útbúa viðmið eða staðla er þá safnað með kerfisbundnu úrtaki sem endurspeglar markhópin. Þessi viðmið eru forsenda þess að unnt sé að leggja mat á stöðu einstaklinga og greina styrk- og veikleika þeirra og eru þau nauðsynleg í allri greiningarvinnu (Crocker og Algina, 1986; Angoff, 1971). Mikilvægt er að skilgreina vandlega markhóp og að draga fram úrtak sem endurspeglar hann sem allra best. Allajafna er eðlilegt að fylgja sömu útfærslum á fyrirlögn, stigakvörðum, aldursbilum og úrtaki og gert er í frumútgáfu. Þó er nauðsynlegt að huga að því hvort eitthvað stangist á við aðstæður hér. Má þar benda á atriði eins og fyrir hvaða undirhópa viðmið eru nauðsynleg, breidd aldursbila og tengingu mælitalna við túlkun. Stöðlunarferlið er mjög sérhæfður hluti próffræðivinnu sem einungis á við um sum mælitæki og því ekki vikið að honum frekar hér heldur bent á ítarlega umfjöllun hjá Peterson, Kolen og Hoover (1989) og Angoff (1971).

Útgáfa

Þessi hluti ferlisins felst í að gera staðfærða þýðingu mælitækis og upplýsingar um hana aðgengilegar hugsanlegum notendum, rannsakendum eða starfandi fagaðilum

(International Test Commission, 2001). Markaðssetning getur falist í formlegri útgáfu og sölu, grein í fag tímariti eða skýrslu sem unnt er að dreifa til notenda eftir þörfum. Þó þýðing og staðfærsla sé ekki gefin út er nauðsynlegt að ganga svo frá upplýsingum um hana og eiginleika hennar að aðrir hugsanlegir notendur geti síðar nálgast þær. Nauðsynlegt er að gera rannsóknir á réttmæti staðfærðrar þýðingar til að kanna að hvaða marki próffræðilegir eiginleikar mælitækisins haldist. Því er eðlilegt að líta svo á að réttmætisrannsóknir séu einn hluti þýðingarferlisins og hér eru þær felldar undir útgáfuhluta. Hlutverk þeirra er að sýna fram á að túlkun niðurstaðna standist og að unnt sé að draga meginályktanir mælitækisins af staðfærðri þýðingu þess (International Test Commission, 2001; Messick, 1989). Ávallt er matsatriði hverjar af réttmætisrannsóknnum sem gerðar hafa verið á frumútgáfu mælitækisins eigi einnig við um staðfærða þýðingu. Þó þarf allajafna að kanna greiningarhæfni staðfærðrar þýðingar, að hvaða marki hugsmíð er mæld með sama hætti og í frumútgáfu og að hvaða marki stigakvarði staðfærðrar þýðingar er sambærilegur frumútgáfu (van de Vijver og Leung, 1997; van de Vijver og Poortinga, 2002). Um staðfærða þýðingu gildir, rétt eins og um frumútgáfu mælitækja, að niðurstöður réttmætisrannsókna afmarka réttmætar túlkanir á niðurstöðum þess og um leið notagildi þess (Messick, 1989).

Lokaorð

Þýðing og staðfærsla mælitækja sem byggjast á tungumáli er flókið ferli. Til að vel takist til við það verður að velja mælitæki af kostgæfni, skipuleggja vinnu við þýðingu í samræmi við aðstæður og byggja endurbætur á upplýsingum úr ólíkum áttum og kynna eiginleika staðfærðrar þýðingar mælitækisins á videigandi hátt. Nauðsynlegt er að vinnuferlið sem heild og forsendur verkþátta séu ljósar þegar ákvarðanir um val á mælitæki, vinnuferli og aðferðir eru teknar. Hér hefur verið fjallað um þýðingu og staðfærslu sem samfelld ferli er hefst þegar þörf fyrir mælitæki verður ljós og lýkur

þegar staðfærð þýðing mælitækis hefur verið gerð aðgengileg. Í of mörgum tilvikum leiða vanhugsaðar eða rangar ákvarðanir til þess að illa tekst til við verkið. Oftast er þetta vegna þess að þeir sem ákvarðanir taka hafa ekki yfirsýn á verkið eða þekkja ekki til þeirra leiða sem færar eru í hverjum verkhluta. Þessari grein er ætlað að bæta þar úr og er hún ætluð rannsakendum, fræðimönnum og fagfólki er notar mælitæki af þessu tagi í félagsvísindum, hugvísindum, menntageiranum, heilbrigðisgeiranum, á viðskiptasviðinu og víðar.

Í þessari grein hefur áherslan verið á að lýsa próffræðilegum forsendum þýðingar og staðfærslu og heildstæðu vinnuferli við verkið. Æskilegt hefði verið að fjalla ítarlegar um verkþætti og aðferðir en unnt var að gera hér. Þörf er á að fjalla á íslensku um ólíkar aðferðir þáttgreiningar til að kanna innri byggingu mælitækja, aðferðir til að kanna atriðabundinn hópamun, stöðlun mælitækja, um atriði sem huga þarf að við val á mælitækjum til þýðingar og fleira. Þau verkefni verða að bíða um hríð.

Abstract

Translation and adaptation of measurement instruments

The purpose of translation/adaptation of measurement instruments to another language is to obtain an instrument that can measure the same construct in identical manner in the second culture. The article presents a psychometric model, based on classical test theory, which describes the necessary conditions for the translation/adaptation to achieve this goal. Secondly, the article describes the translation/adaptation process as a five component work process consisting of preparation, translation, revisions, standardization (if relevant), and publication. The preparation part consist of all actions up to the point where decision is made to translate/adapt. Translation consist of translating the instrument text and adapting it to the new culture. Revisions consist of evaluating the quality of translated items and revising items that are judged to be flawed.

Standardization consists of defining and sampling from the target population, collecting data and constructing norms. The publication process includes conducting validity studies and publishing information regarding the translation/adaptation appropriately. An overall view is gained by presenting translation/adaptation in this manner, a view that will simplify, particularly for individuals with little experience in measurement or regarding translation/adaptation, the planning and preparation for translation/adaptation projects.

Heimildir

- Allouf, A., Hambleton, R. K. og Sireci, S. G. (1999). Detecting the causes of differential item functioning in translated verbal items. *Journal of Educational Measurement*, 36, 185–198.
- Angoff, W. H. (1971). Scales, norms and equivalent scores. Í R. L. Thorndike (Ritstj.), *Educational measurement* (2. útg., bls. 508–600). Washington, DC: American Council on Education.
- Angoff, W. H. (1982). Use of difficulty and discrimination indices for detecting item bias. Í R. A. Berk (Ritstj.), *Handbook for methods of detecting test bias* (bls. 96–116). Baltimore, MD: Johns Hopkins University Press.
- Arce-Ferrer, A. J. (2005). *Investigating translation and adaptation of rating scales for Spanish-speaking populations*. Erindi flutt á ráðstefnunni Annual meeting of the National Council on Measurement in Education, Montreal, Kanada, 12.–15. apríl 2005.
- Baumgartner, H. og Steenkamp, J. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, 28, 143–156.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Brachen, B. A. og Barona, A. (1991). State of the art procedures for translating, validating and using psychological tests in cross-cultural assessment. *School Psychology International*, 12, 119–132.
- Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-cultural Psychology*, 1, 185–216.
- Brislin, R. W. (1986). The wording and translation of research instruments. Í W. J. Lonnes og J. W. Berry (Ritstj.), *Field methods in cross cultural research* (bls. 137–164). Beverly Hills, CA: Sage Publications.
- Clauser, B. og Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues & Practice*, 17(1), 31–44.
- Crocker, L. og Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando, FL: Harcourt Brace.
- Dorans, N. J. (1989). Two new procedures to assess differential item functioning: The standardization method and the Mantel-Haenzel method. *Applied Measurement in Education*, 2, 217–233.
- Ellis, B. B. (1989). Differential item functioning: Implications for test translations. *Journal of Applied Psychology*, 74, 912–920.
- Ercikan, K. (1998). Translation effects in international assessments. *International Journal of Educational Research*, 29, 543–553.
- Feldt, L. S. og Brennan, R. L. (1989). Reliability. Í R. L. Linn (Ritstj.), *Educational measurement* (3. útg., bls. 105–146). Phoenix, AZ: American Council on Education, Oryx Press.

- Fowler, F. J. (2000). *Improving Survey Questions: Design and Evaluation*. Thousand Oaks, CA: Sage Publications.
- Gierl, M. J. (2000). Construct equivalence on translated achievement tests. *Canadian Journal of Education*, 25, 280–295.
- Gierl, M. J. og Khaliq, S. N. (2001). Identifying sources of differential item functioning and bundle functioning on translated achievement tests: A confirmatory analysis. *Journal of Educational Measurement*, 38, 164–187.
- Grisay, A. (2003). Translation and procedures in the OECD/PISA 2000 international assessment. *Language Testing*, 22, 225–240.
- Grisay, A. (2004). Translation and cultural appropriateness of the test and survey material. Í R. Adams og M. Wu (Ritstj.), *PISA 2000 technical report* (bls. 57–70). OECD: Paris.
- Gulliksen, H. (1950/1987). *Theories of mental tests*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hambleton, R. K. (1993). Translating achievement tests for use in cross-national studies. *European Journal of Psychological Assessment*, 9, 57–68.
- Hambleton, R. K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. Í R. K. Hambleton, P. F. Merenda og C. D. Spielberger (Ritstj.), *Adapting educational and psychological tests for cross-cultural assessment* (bls. 3–38). Mahwah, N.J.: Lawrence Erlbaum Associates.
- Hambleton, R. K. og Bollwark, J. (1991). Adapting tests for use in different cultures: Technical issues and methods. *International Test Bulletin*, 32/33, 3–32.
- Heimir Pálsson og Höskuldur Þráinsson. (1988). *Um þýðingar*. Reykjavík: Íðunn.
- Hulin, C. L. (1987). A psychometric theory of evaluations of item and scale translations: Fidelity across languages. *Journal of Cross-Cultural Psychology*, 18, 115–142.
- International Test Commission (2001). *ITC test adaptation guidelines*. Sótt 25. október, 2001 af http://www.intestcom.org/test_adaptation.htm
- Krueger, R. A. og Casey, M. A. (2000). *Focus groups: A practical guide for applied research* (3.útg). Thousand Oaks, CA: Sage Publication.
- Maxwell, B. (1996). Translation and cultural adaptation of the survey instruments. Í M. O. Martin og D. L. Kelly (Ritstj.), *Third international mathematics and science study (TIMSS) technical report, Volume I: Design and development* (bls. 8.1–8.10). Chestnut Hill, MA: Boston College.
- McDonald, (1985). *Factor analysis and related methods*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Messick, S. (1989). Validity. Í R. L. Linn (Ritstj.), *Educational measurement* (3. útg., bls. 13–103). Phoenix, AZ: American Council on Education, Oryx Press.
- Millman, J. og Greene, J. (1989). The specification and development of tests of achievement and ability. Í R. L. Linn (Ritstj.), *Educational measurement* (3. útg., bls. 335–366). Phoenix, AZ: American Council on Education, Oryx Press.
- Peterson, N. S., Kolen, M. J. og Hoover, H. D. (1989). Norms, scales, and equivalent scores. In R. L. Linn (Ed.) *Educational measurement* (3. útg., bls. 221–262). Phoenix, AZ: American Council on Education, Oryx Press.

- Prieto, A. J. (1992). A method for translation of instruments to other languages. *Adult Education Quarterly*, 43, 1–14.
- Sigurgrímur Skúlason (2004a). *An investigation of the validity of the Iowa Early Learning Inventory*. Óbirt doktorsritgerð: University of Iowa.
- Sigurgrímur Skúlason (2004b). *Hegðun 5 og 6 ára barna sem tengist námi: Pýðing og staðfærsla á matslista*. Erindi flutt á ráðstefnunni Rannsóknir – Nýbreytni – Próun: 8. málþing Rannsóknastofnunar Kennaraháskóla Íslands, Reykjavík, 16. október 2004.
- Sireci, S. G., Patsula L. og Hambleton, R. K. (2005). Statistical methods for identifying flaws in the test adaptation process. Í R. K. Hambleton, P. F. Merenda og C. D. Spielberger (Ritstj.), *Adapting educational and psychological tests for cross-cultural assessment* (93–115). Mahwah, N.J: Lawrence Erlbaum Associates.
- Solano-Flores, G., Contreras-Niño, L. Á. og Backhoff, E. (2005). *The Mexico translation of TIMSS-1995: Lessons on test translation from a post-mortem study*. Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, Kanada, 12–15 apríl 2005.
- Tanzer, N. K. (1995). Cross-cultural validity of Likert-type scales: Perfect matching factor structures and still biased. *European Journal of Psychological Assessment*, 11, 194–201.
- Thissen, D., Steinberg, L. og Gerrard, M. (1986). Beyond group-mean differences: The concept of item bias. *Psychological Bulletin*, 99, 118–128.
- van de Vijver, F. og Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist*, 1, 89–99.
- van de Vijver, F. og Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage Publications.
- van de Vijver, F. og Poortinga, Y. H. (1997). Towards integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment*, 13, 29–37.
- van de Vijver, F. og Poortinga, Y. H. (2002). Structural equivalence in multilevel research. *Journal of Cross-cultural Psychology*, 33, 141–156.
- van de Vijver, F. og Poortinga, Y. H. (2005). Conceptual and methodological issues in adapting tests. Í R. K. Hambleton, P. F. Merenda og C. D. Spielberger (Ritstj.), *Adapting educational and psychological tests for cross-cultural assessment* (39–63). Mahwah, N.J: Lawrence Erlbaum Associates.
- van Herk, H., Poortinga, Y. H. og Verhallen, T. M. M. (2004). Response styles in rating scales: Evidence of bias in data from six EU countries. *Journal of Cross-cultural Psychology*, 35, 346–360.
- Wong, N., Rindfleisch, A. og Burroughs, J. E. (2003). Do reverse-worded items confound measures in cross-cultural consumer research? The case of materials values scale. *Journal of Consumer Research*, 30, 72–91.
- Þorlákur Karlson (2003). Spurningakannanir: Uppbygging, orðalag og hættur. Í Sigríður Halldórsdóttir og Kristján Kristjánsson (Ritstj.), *Handbók í aðferðafræði og rannsóknum í heilbrigðisvísindum* (bls. 331–355). Akureyri: Háskólinn á Akureyri.