Frank Bach är verksam vid institutionen för didaktik och pedagogisk profession vid Göteborgs universitet och ingår i den grupp som, för Skolverkets räkning, har utformat nationella prov i NO-ämnena för årskurs 6. Hans utbildnings- och forskningsintresse är ämnesdidaktik med inriktning mot naturvetenskap och framförallt då fysik.

Birgitta Frändberg är verksam vid institutionen för didaktik och pedagogisk profession vid Göteborgs universitet och ingår i den grupp som, för Skolverkets räkning, har utformat nationella prov i NO- ämnena för årskurs 6. Hennes utbildnings- och forskningsintresse är ämnesdidaktik med inriktning mot naturvetenskap, framförallt biologi och kemi.

Mats Hagman är verksam vid institutionen för didaktik och pedagogisk profession vid Göteborgs universitet och ingår i den grupp som, för Skolverkets räkning, har utformat nationella prov i NO-ämnena för årskurs 6. Hans utbildnings- och forskningsintresse är ämnesdidaktik med inriktning mot naturvetenskap.

Eva West är verksam vid institutionen för didaktik och pedagogisk profession vid Göteborgs universitet och ingår i den grupp som, för Skolverkets räkning, har utformat nationella prov i NO-ämnena för årskurs 6. Hennes utbildnings- och forskningsintresse är ämnesdidaktik med inriktning mot naturvetenskap.

Ann Zetterqvist är verksam vid institutionen för didaktik och pedagogisk profession vid Göteborgs universitet och ingår i den grupp som, för Skolverkets räkning, har utformat nationella prov i NO-ämnena för årskurs 6. Hennes utbildnings- och forskningsintresse är ämnesdidaktik med inriktning mot naturvetenskap, framförallt biologi och kemi.

FRANK BACH
Göteborgs universitet, Sverige
frank.bach@gu.se

BIRGITTA FRÄNDBERG
Göteborgs universitet, Sverige
birgitta.frandberg@ped.gu.se

MATS HAGMAN
Göteborgs universitet, Sverige
mats.hagman@ped.gu.se

EVA WEST
Göteborgs universitet, Sverige
eva.west@ped.gu.se

ANN ZETTERQVIST
Göteborgs universitet, Sverige
ann.zetterqvist@ped.gu.se

Att betygsätta elevers kunskaper – möjligheter och begränsningar

Abstract

National tests in the natural sciences for grade 6 were carried out 2013 - 2015. One aim was to provide information about students' scientific knowledge according to the "abilities": communicating, exploring and explaining, as stated in the syllabus. Are these "abilities" possible and reasonable to distinguish from each other? Answers (60,000) were used for exploratory and confirmatory factor analyses and controlled for dimensionality in a Rasch model. The results show that the tests mainly capture one overall ability. It was not possible to separate the three "abilities" in a reliable way. Thus, our analyses show that dividing the test into subscales for grading the "abilities" jeopardize test validity. A more reliable option might be to let students' strong and weak performances in relation to different parts of the syllabus compensate for each other. Lastly we discuss our results in relation to the Swedish educational legislation.

Inledning

I denna artikel avser vi att diskutera de problem med betygsättning som kan uppkomma när man försöker dela upp kunskap i ett ämne i flera mindre delar. Med hjälp av statistiska analyser av nationella prov i de naturorienterande ämnena (NO-ämnena) undersöker vi om olika dimensioner i kunskaper kan urskiljas på ett tillförlitligt sätt och hur reliabiliteten påverkas om delprovsbetyg skulle sättas. Detta diskuteras i förhållande till de formuleringar som reglerar betygsättningen i de svenska styrdokumenten och till den praxis som har utvecklats vid betygsättning i NO-ämnena.

Historisk utveckling av NO-kursplanerna

Formuleringarna i de svenska kursplanerna i NO-ämnena i grundskolan har historiskt sett påverkats av två internationella trender. Den ena, som är generell och gäller alla skolans ämnen, rör frågan om vilka kunskapsuttryck som anses giltiga och den andra handlar om innehållsliga förändringar i NO-ämnena.

Beträffande frågan om kunskapsuttryck uppvisar den internationella diskussionen kring kunskaper, läroplaner och utvärdering en förändring från att *kunna* till att *kunna göra* (Yates & Collins, 2010). Detta gäller även synen på vad det innebär att lära sig naturvetenskap, som har förändrats från att handla om individens förståelse och kunnande till deltagande i en praktik (Kind, 2013). När det gäller bedömning av elevers kunskaper fokuseras i den senaste svenska läroplanen (i kunskapskraven) vad eleverna ska kunna göra. Detta till skillnad från tidigare läroplaner, som i högre grad uppehöll sig vid vilka kunskaper eleverna skulle utveckla, vilket stämmer väl med förloppen i andra länder (se t.ex. Yates & Collins, 2010).

De innehållsliga förändringarna om vad som anses vara viktigt i naturvetenskap innebar att fokus i läroplaner under 1950- och 1960-talen flyttades från att ha varit begreppsinriktade till att tydligare lyfta fram naturvetenskapens karaktär och processer (Orpwood, 2001, 2007). Intresset riktades mot hur naturvetenskaperna fungerar som discipliner och de färdighetskunskaper som präglar det som kallades ”det naturvetenskapliga arbetssättet”. Denna syn kritiserades under senare delen av 1960-talet och under 1970-talet eftersom den endast ansågs vara relevant för ett mindre antal elever, nämligen de som planerade en fortsatt utbildning inom naturvetenskap och teknik (Stuckey, Hof-

stein, Mamlok-Naaman, & Eilks, 2013). De diskussioner som sedan följde och som starkt influerade förändringar i läroplaner kom bland annat att handla om ”Scientific Literacy for All”. Det betyder att om det första skiftet handlar om att blicka inåt i naturvetenskaperna och deras metoder för att utveckla kunskap vänder sig det andra skiftet utåt mot relationerna mellan naturvetenskap, teknik och samhällsutveckling. Genom detta senare skifte framträder tydligt frågor om naturvetenskapens roll vid demokratiska beslut och personliga val (Orpwood, 2001, 2007).

Tecken på de ovan beskrivna synsätten är tydliga i dagens svenska kursplaner för NO-ämnena. Exempelvis sammanfattas syftet med undervisningen i tre punktsatser som i tur och ordning lyfter fram samhällskopplingen, de kunskapsgenererande metoderna och naturvetenskapernas begrepp, modeller och teorier. För ämnet kemi är denna sammanfattning formulerad på följande sätt (Skolverket, 2018, s. 185):

Genom undervisningen i ämnet kemi ska eleverna sammanfattningsvis ges förutsättningar att utveckla sin förmåga att

- *använda kunskaper i kemi för att granska information, kommunicera och ta ställning i frågor som rör energi, miljö, hälsa och samhälle,*
- *genomföra systematiska undersökningar i kemi, och*
- *använda kemins begrepp, modeller och teorier för att beskriva och förklara kemiska samband i samhället, naturen och inuti människan.*

De tre punktsatserna ovan är ett slags långsiktiga mål för undervisningen. De uttrycker vad undervisningen ska bidra till, men tolkas i många sammanhang som att det handlar om tre distinkt urskiljbara ”förmågor” som eleverna ska tillägna sig, och där varje ”förmåga” kan och bör prövas separat. Det faktum att kunskapskravens olika formuleringar är direkt kopplade till de tre punktsatserna bidrar till denna tolkning (Skolverket, 2018).

Fortsättningsvis kommer de ovan beskrivna punktsatserna att för enkelhetens skull benämnas förmågan att kommunicera, förmågan att undersöka respektive förmågan att använda begrepp.

Det svenska betygssystemet

Det svenska betygssystemet är målrelaterat och har en betygsskala med sex steg: A, B, C, D, E och F. Det finns preciserade kunskapskrav som anger vilka kunskaper som krävs för betygen A, C och E men inte för de mellanliggande betygen. Betygssystemet har beskrivits som icke-kompensatoriskt, eftersom det tolkats som att goda prestationer inom ett område inte kan uppväga mindre goda prestationer inom ett annat (Gustafsson, Cliffordsson, & Erickson, 2014). Elever har också vittnat om att det i vissa fall varit oerhört svårt att nå de högsta betygen då det uppfattats som att den sämsta prestationen påverkar betyget mer än den bästa prestationen. I en utvärdering av Skolverket (Skolverket, 2016b) pekas den så kallade ”tröskelregeln” ut som ett särskilt problem, det vill säga formuleringarna om att kunskapskraven ska vara uppfyllda i sin helhet för ett visst betyg. Det har också spridit sig en praxis bland lärare att vid bedömning och betygsättning dela upp kunskapskraven i mindre delar och att bedöma varje del för sig. Ursprunget till uppfattningen att man bör dela upp kunskapskraven i delar, och att det är de sämsta prestationerna som betyder mest, står förmodligen att finna i de formuleringar som reglerar betygsättningen, exempelvis nedanstående formulering i Skolförordningen (SFS 2011:185, s. 506):

> *Kraven för betygen A, C och E ska precisera vilka kunskaper som krävs för respektive betyg. Kunskapskravet för betyget D innebär att kraven för E och till övervägande delen för C är uppfyllda. Kravet för betyget B innebär att även kraven för C och till övervägande delen för A är uppfyllda.*

Hur lärare tolkar vad som är en del av kunskapskravet varierar mellan att hela beskrivningen av en nivå, exempelvis E-nivån, utgör kunskapskravet, till att varje mening eller del av mening där det finns så kallade värdeord är en del av kunskapskravet. Den sistnämnda tolkningen kan ge upphov till att kunskapskravet för varje NO-ämne i årskurs 6 byggs upp av minst 12 kravformuleringar. En vanlig tolkning är att varje kunskapskrav består av tre delar, nämligen de tre ”förmågor” som beskrivits ovan. De nationella proven i NO för årskurs 6 konstruerades utifrån en idé om dessa tre ”förmågor”.

Tolkningen av vad styrdokumenten säger om bedömning, till exempel hur kunskapskraven ska förstås, har sedan flera år varit föremål för debatt i Sverige. Det som lärare upplever som mest problematiskt är hur man ska bedöma elever med ojämn kunskapsprofil i det icke-kompensatoriska betygssystemet (Skolverket, 2016b, s. 73), och regeringen har meddelat att man avser utreda om betygssystemet även i fortsättningen ska vara icke-kompensatoriskt (Regeringen, 2017, s. 27-28).

Nationella prov i naturorienterande ämnen för årskurs 6

Det främsta syftet med de svenska nationella proven är att stödja en likvärdig och rättvis bedömning och betygssättning. De är i princip konstruerade för att testa hur väl eleverna uppfyller de uppsatta kunskapskraven. Eleverna får ett provbetyg på det prov de deltagit i, och det används, tillsammans med annan information läraren har om elevens kunskaper, som ett underlag när betyg sätts.

Nationella prov i NO-ämnena för årskurs sex genomfördes under tre år; 2013 och 2014 då de var obligatoriska samt 2015 då varje skola kunde bestämma om de ville använda dem eller inte. På varje deltagande skola gavs prov i ett av ämnena biologi, fysik och kemi. För alla tre provomgångarna gällde att lärarna skulle bedöma elevernas svar utifrån bedömningsanvisningar som tillhandahölls av Skolverket.

Varje ämnesprov bestod av tre delprov för att mäta elevers kunskaper i enlighet med de tre punktsatserna, ”förmågorna”, i sammanfattningen av syftestexten i kursplanerna. På en webbsida (http://npno6.se/page/exempelprov.php[1]) som Skolverket länkade till beskrevs delproven på följande kortfattade sätt:

Ämnesproven i NO-ämnena för årskurs 6 kommer att vara uppdelade i tre delprov, ett för varje förmåga (se Om proven)

Delprov A: Förmågan att kommunicera
Delprov B: Förmågan att undersöka
Delprov C: Förmågan att använda begrepp

Varje elev genomförde prov i ett av NO-ämnena och besvarade då uppgifter i tre delprov, som vart och ett fick ta en timme att genomföra. Delproven hade olika karaktär, inte bara avseende vilken typ av kunskap de avsåg att testa utan också i sådant som antal uppgifter och uppgiftsformat. Delprov A i alla ämnen omfattade tre till fyra uppgifter som vanligen innehöll mycket text att läsa och som krävde relativt omfattande svar. Delprov B omfattade drygt dubbelt så många uppgifter med lite mindre mängd text att läsa och skriva och innehöll vanligen också en praktisk uppgift. I Delprov C var det knappt 20 uppgifter, med en blandning av flervalsfrågor och uppgifter som krävde i huvudsak kortare egenformulerade svar. Varje deltagande elev fick ett provbetyg som beräknades genom en viktad summering av resultaten på de tre delproven. Viktningen var konstruerad så att varje delprov skulle väga lika mycket oavsett hur många poäng man kunde få på dem. Uppgifterna från proven i biologi, fysik och kemi finns tillgängliga i sin helhet och kan nås via Skolverkets hemsida.

[1] Denna webbplats finns inte längre, men motsvarande information går att erhålla via t.ex. https://bp.skolverket.se/delegate/download/test/informationmaterial?testGuid=E73DA0B5DF454A199053C646837377781

Mätsäkerhet i prov

Det gängse mätteoretiska antagandet när det gäller bedömning av elevers kunskaper är att det finns ett oändligt antal uppgifter för att pröva elevers kunnande inom ett specifikt område. På liknande sätt som slumpmässiga urval av populationer av människor hanteras för att uttala sig om populationerna som helhet kan de uppgifter som används för att pröva elevers kunnande hanteras som ett slumpmässigt urval av ett oändligt antal möjliga indikationer på elevers kunskapsnivå. Ett fåtal indikationer (t.ex. resultat från få uppgifter) innebär större osäkerhet när det gäller en elevs kunskapsnivå på samma sätt som ett fåtal individer innebär större osäkerhet när det gäller att uttala sig om populationen som helhet. I ett prov kommer en elev alltid att möta ett urval av ett, teoretiskt sett, oändligt antal uppgifter.

Den praktiska konsekvensen är att man behöver många uppgifter för att kunna uttala sig om en elevs kunskapsnivå inom ett visst område. Delar man upp prov i olika delar, till exempel utifrån idén om olika förmågor, riskerar man att tappa i reliabilitet, om man inte kompenserar uppdelningen med att öka antalet uppgifter (se t.ex. Gustafsson, 2008, s. 5).

Dimensionalitet i mätningar

En vanlig utgångspunkt vid provkonstruktion är att man mäter en underliggande egenskap hos respondenterna, till exempel en elevs biologikunskaper. I praktiken är detta en approximation eftersom fullständig endimensionalitet i mätning av kunskap inte är möjlig om man har fler än en mätpunkt, exempelvis fler än en uppgift i ett prov. Uppgifter som är tänkta att pröva elevers kunskaper i till exempel fotosyntes eller i evolutionsbiologi mäter olika aspekter av biologikunskaper och en elev kan vara mycket kunnig i den ena men mindre kunnig i den andra. Kunskaperna som mäts kan ändå betraktas som tillräckligt endimensionella om det är möjligt att göra meningsfulla tolkningar av mätningen, i det här fallet som ”biologikunskaper” (se t.ex. Sjaastad, 2014, s. 214). Eftersom elevernas kunskaper ska betygsättas utefter en endimensionell betygsskala behöver man betrakta kunskaper i skolämnen som endimensionella.

Syfte och forskningsfrågor

Med utgångspunkt i diskussionen om icke-kompensatorisk betygssättning och med regeringens planerade utredning i fråga om icke-kompensatoriska eller kompensatoriska betyg i åtanke (Regeringen, 2017, s. 27–28), vill vi med denna studie undersöka om det är möjligt att urskilja de delar av NO-kunskap, ”förmågor”, som man påstår sig mäta i de nationella proven. Resultatet av undersökningen används som grund för att diskutera huruvida det är rimligt att betrakta ”förmågorna” i kursplanerna som separata storheter. Handlar det om olika dimensioner i kunskapsbegreppet, var och en så väsentlig att den ensam bör avgöra högsta möjliga nivå på en elevs betyg eller handlar det om olika aspekter i ett sammanvävt kunskapsbegrepp där en elevs starka sidor kan kompensera för de svagare? Underlaget för studien är elevresultat från nio nationella prov i NO-ämnen där ambitionen har varit att konstruera uppgifter i syfte att pröva eleverna mot de tre angivna ”förmågorna”.

- Går det att, på ett reliabelt sätt, urskilja de så kallade ”förmågorna” från varandra i elevernas svar på de nationella proven i NO-ämnena för årskurs 6?
- Hur påverkas reliabiliteten av en uppdelning av provet i mindre delar?

Metod

För att besvara forskningsfrågorna genomfördes statistiska analyser av kvantitativa data. Detta kompletterades med kvalitativa analyser av uppgifters struktur och innehåll med avseende på de krav som ställdes på elevernas kunskaper i nio nationella prov mellan 2013 och 2015.

Dataunderlag

Data utgörs av mer än 60 000 elevers resultat på uppgifter i nationella prov i biologi, fysik och kemi för årskurs 6, åren 2013, 2014 och 2015.

Den information som användes i analyserna har genererats genom att lärare frivilligt lagt in sina elevers resultat på provens uppgifter på en webbplats, mot att de fått hjälp att beräkna det sammanvägda provbetyget för eleverna. Vi har därigenom fått tillgång till resultat på enskilda provfrågor för ett mycket stort antal elever. För åren 2013 och 2014, när de nationella proven i NO-ämnena var obligatoriska, har Skolverket samlat in och publicerat statistik över bland annat totalt antal deltagande elever. Vårt urval utgör för dessa år 32% respektive 25% av det totala antalet elever som genomförde proven (Tabell 1). För år 2015, när proven var frivilliga och många skolor valde att inte genomföra dem, finns ingen sådan statistik publicerad av Skolverket.

Tabell 1. Antal elever i vårt urval i förhållande till det totala antalet elever som gjort proven.

	Antal elever		
År	Vårt urval	Totalt	Andel (%)
2013	29 931	92 473	32,4
2014	23 138	93 712	24,7
2015	10 298	-	-

Då informationen inte genererades genom ett slumpmässigt urval så undersöktes i vilken utsträckning data är representativa för hela de populationer som genomförde de nationella proven i biologi, fysik och kemi. Detta gjordes genom en jämförelse av fördelningen av provbetyg i våra urval med data från Skolverkets totalinsamling av sammanvägda provbetyg för de nationella proven.

Skillnaderna i fördelningen av provbetyg mellan våra urval och Skolverkets totaldata är små. I en tidigare studie har representativiteten för vårt urval året 2013 analyserats mer i detalj (Bach, Frändberg, Hagman, West, & Zetterqvist, 2015), bland annat avseende fördelning mellan flickor och pojkar samt andelen elever med svenska som andraspråk. Skillnaderna gentemot Skolverkets totaldata var små även i dessa avseenden och slutsatsen blev att representativiteten är god. För syftet med denna artikel är det viktigt att konstatera att variabiliteten i data inte heller avviker i någon avgörande mening från Skolverkets totalinsamling.

Analyser

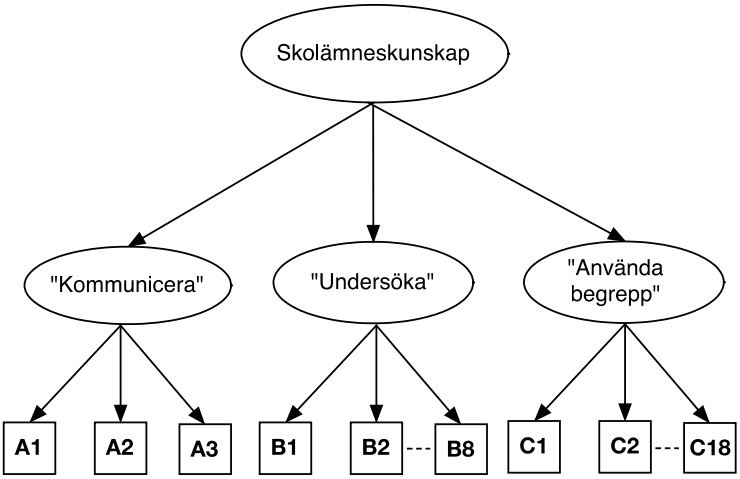
De data som har analyserats består av lärares poängsättning av elevsvar på de enskilda uppgifterna. Varje ämnesprov är som tidigare nämnts uppdelat i tre delprov där delprov A har utvecklats för att pröva elevernas förmåga att ”kommunicera”, delprov B för elevernas förmåga att ”undersöka” och delprov C elevernas förmåga att ”använda begrepp”.

Studiens första forskningsfråga berör hur många bakomliggande faktorer som är rimligt att urskilja för att förklara elevers resultat på de nationella proven i de tre NO-ämnena. Går det att på ett tillförlitligt sätt urskilja de tre ”förmågorna” (”kommunicera”, ”undersöka” respektive ”använda begrepp”) i datamaterialet, eller är det rimligare att betrakta resultaten som ett uttryck för en enda överordnad dimension?

Tre olika metoder har använts för att undersöka i vilken utsträckning de tre “förmågorna” framträder empiriskt:

- Explorativ faktoranalys (Mplus, ordinalskala, (Muthén & Muthén, 2016)
- Konfirmatorisk faktoranalys (Mplus, ordinalskala, (Muthén & Muthén, 2016)
- Kontroll av dimensionalitet i data i förhållande till Rasch-modeller av proven

Explorativ faktoranalys är en teknik som går ut på att bestämma det minsta rimliga antalet faktorer som beskriver sambanden mellan ett antal observerade variabler. Metoden är utforskande i den meningen att den inte förutsätter en struktur för förhållandet mellan de observerade variablerna och de bakomliggande faktorerna. Konfirmatorisk faktoranalys utgår, till skillnad från den explorativa, från i förväg teoretiskt antagna bakomliggande faktorer som kan förklara mönster i de data man undersöker (Pedhazur & Pedhazur Schmelkin, 1991). I vårt fall utgörs de tre faktorerna av det som omnämns som "förmågor" i kursplanerna för NO-ämnena i grundskolan. Elevernas resultat från de tre delproven definieras som uttryck för varsin faktor. Utöver dessa faktorer definieras också en övergripande faktor, skolämneskunskap, som påverkar de tre "förmågorna" (Brown, 2015, s. 287–288) här (se Figur 1). Med termen skolämneskunskap avses en elevs kunskaper i ett av skolämnena biologi, fysik och kemi. Tekniken att på detta sätt fastställa så kallade latenta variabler har bland annat den fördelen att de är fria från mätfel (Gustafsson, 2009, s. 269).



Figur 1. Illustration av den konfirmatoriska faktormodellen. A1-A3, B1-B8, C1-C18 representerar resultat från de olika provuppgifterna.

I exemplet relaterar uppgifterna A1-A3 endast till faktorn "Kommunicera" och inte till de andra faktorerna. Motsvarande gäller för faktorerna "Undersöka" och "Använda begrepp". Genom att faktorerna (de latenta variablerna) korrelerar med varandra, är det möjligt att definiera en latent variabel i andra ordningen, "skolämneskunskap", i vart och ett av de tre skolämnena. Pilarna i Figur 1 indikerar att den överordnade faktorn, "skolämneskunskap", antas påverka elevens förmågor att "kommunicera", "undersöka" respektive "använda begrepp". Dessa "förmågor" påverkar i sin tur elevens resultat på de enskilda uppgifterna i respektive delprov.

En mätteknisk utmaning är att data från proven är på ordinalskala och maximalt antal poäng på uppgifterna varierar mellan ett och tre. Detta ställer specifika krav på de mjukvaror som kan användas. Vi har använt Mplus som har rutiner för både explorativ och konfirmatorisk faktoranalys för ordinalskala (Muthén & Muthén, 2016).

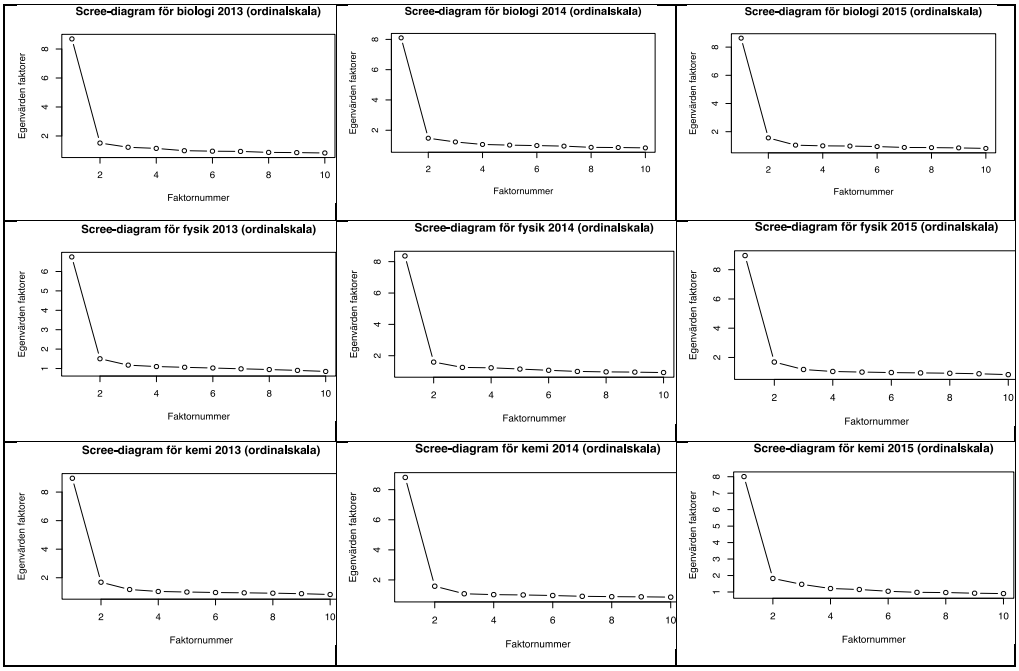
Ett annat sätt att hantera ordinalskala för data från uppgifterna är att använda Item Response Theory som till exempel applicerar en endimensionell modell genom att med iterativa metoder transformera data på ordinal- till intervallskala. I studien används Rasch-modellering (Bond & Fox, 2001) för att undersöka provens dimensionalitet. Vi använder standardiserade metoder för Rasch-modellen för att undersöka om uppgifterna i proven passar till modellen och principalkomponentanalys (Chou & Wang, 2010; Sjaastad, 2014) av residualerna för att fånga eventuella dimensioner utöver raschmodellens. Detta görs med hjälp av modulen Pairwise (Heine, 2017) i statistikprogrammet R (R Development Core Team, 2008).

Den andra forskningsfrågan handlar om hur reliabiliteten påverkas om provet delas upp i mindre delar. Vi har använt delar av vårt datamaterial för att undersöka vilka konsekvenserna skulle bli av ett sådant förfarande. Eftersom delproven A omfattar minst antal uppgifter och därmed har störst risk för låg reliabilitet är de intressanta att undersöka beträffande vilka konsekvenser en icke-kompensatorisk uppdelning skulle få för provbetyget. Ett sätt att analysera konsekvenserna är att utgå från beräkningar av reliabilitet hos delprov A i de nio prov som gavs under åren 2013-2015 för att beräkna medelfelet i mätningen (SEM). Detta beräknas enligt $SEM = s\sqrt{1-r}$ där s betecknar standardavvikelsen för den observerade poängen på provet och r betecknar reliabiliteten. Den enskilda provtagarens sanna poäng x_s ligger då med 95 procents sannolikhet mellan x-1,96•*SEM*≤ x_s≤x+1,96•*SEM* där x står för den uppnådda poängen på provet (Crocker & Algina, 2008, s. 122–123).

Resultat

Explorativ faktoranalys

I ett första steg i den explorativa faktoranalysen konstrueras så kallade scree-diagram (se Figur 2). Sådana diagram visar hur stor betydelse olika bakomliggande faktorer har för att beskriva strukturen i de data man analyserar. I ett scree-diagram undersöks om det finns en eller flera faktorer med höga



Figur 2. Scree-diagram för nio nationella prov i NO-ämnen för årskurs 6 åren 2013-2015

värden. Varje faktor som i ett sådant diagram uppvisar stor skillnad mot nästkommande faktor tas med i en fortsatt analys. Resten av faktorerna bedöms erbjuda för litet förklaringsvärde för att vara av intresse. Ett annat sätt att uttrycka detta är att de faktorer som i stort sett bildar punkter på en rät linje inte tas med.

Scree-diagrammen uppvisar ett liknande utseende för alla nio proven. En tydlig faktor framträder och därutöver en mindre tydlig. Möjligen kan man diskutera en tredje faktor, särskilt i kemiprovet 2015. Då proven är utformade med tre olika delar som syftar till att pröva eleverna i förhållande till de tre ”förmågorna” borde tre tydliga faktorer framträda i diagrammen. Så är dock inte fallet.

Nästa steg i analysen är att undersöka hur väl data passar med enfaktormodellen. Detta har prövats mot standardiserade gränsvärden (fit indices) (se Tabell 3). Värdena för CFI och TLI bör vara nära 0,95 eller högre, RMSEA lägre än 0,05 och SRMR lägre än 0,08 för en bra modellpassning (Hu & Bentler, 1999).

Tabell 2. Sammanställning av hur data passar en enfaktormodell för nio nationella prov i NO-ämnen för årskurs 6.

Prov	χ^2	*df*	CFI	TLI	RMSEA	SRMR
Biologi 2013	5096*	434	0,95	0,95	0,034	0,040
Biologi 2014	3581*	377	0,97	0,96	0,032	0,037
Biologi 2015	1660*	405	0,96	0,96	0,031	0,036
Fysik 2013	4844*	350	0,94	0,94	0,034	0,038
Fysik 2014	8684*	560	0,90	0,90	0,043	0,043
Fysik 2015	4103*	434	0,91	0,91	0,049	0,045
Kemi 2013	7890*	405	0,93	0,93	0,044	0,044
Kemi 2014	3856*	464	0,96	0,96	0,032	0,036
Kemi 2015	11347*	464	0,91	0,90	0,083	0,105
Kemi 2015 (-C13)	1685*	405	0,96	0,96	0,030	0,041

Not: χ^2 = chi-square goodness of fit statistic; df = degrees of freedom; CFI = Comparative Fit Index; TLI = Tucker Lewis Index; RMSEA = Root Means-Square Error of Approximation; SRMR = Standardised Square Root Mean Residual. * χ^2 är statistiskt signifikant.

Det enligt Gustafsson (Gustafsson, 2009, s. 293–294) viktigaste måttet, RMSEA, ligger under gränsvärdet 0,05 för alla prov utom kemi 2015. I detta prov finns det en uppgift i Delprov C, C13, som sticker ut och möjligen prövar något annat än vad som var avsett. Genom att ta bort denna uppgift ur analysen passar data bättre till enfaktormodellen. Övriga värden ligger inom eller nära gränsvärdena.

Konfirmatorisk faktoranalys

Så som har beskrivits i metodavsnittet fördes de provuppgifter som har utvecklats för att pröva elevernas kunnande i förhållande till de tre ”förmågorna” till varsin faktor. En övergripande faktor som består av de tre ”förmågefaktorerna” definierades också i modellen. Relationen mellan den överordnande faktorn och de tre underliggande faktorerna redovisas i Tabell 3 som ”laddningar”. ”Laddning” innebär i vilken grad den överordnade latenta variabeln (faktorn) påverkar den underliggande.

Tabell 3. Relationer mellan första och andra ordningens faktorer redovisade som laddningar

Prov	Laddning Förmåga A	Laddning Förmåga B	Laddning Förmåga C
Biologi 2013	0,937	0,937	0,896
Biologi 2014	0,925	0,921	0,923
Biologi 2015	1,031*	0,906	0,921
Fysik 2013	0,866	0,948	0,885
Fysik 2014	0,811	0,964	0,855
Fysik 2015	0,781	0,997	0,888
Kemi 2013	0,839	0,977	0,862
Kemi 2014	0,905	0,922	0,894
Kemi 2015	0,915	0,947	0,877

*I Biologi 2015 tyder laddningen sannolikt på en slumpmässig, icke-signifikant, avvikelse från 1 (ett så kallat Heywood-fall).

De höga värdena i Tabell 3 visar att den överordnade faktorn influerar mycket på de tre underliggande faktorerna. Detta tyder på att den övergripande faktorn, skolämneskunskap, hos eleverna ger utslag i de övriga. Exempelvis slår en ökning i den övergripande faktorn igenom till 93,7% i förmåga A för Biologi 2013.

I Tabell 4 redovisas värden för kvaliteten på denna modell för de nio proven. I samtliga fall är modellpassningen bättre för denna modell än för enfaktormodellen (se Tabell 2). Det kan vara värt att uppmärksamma att fysikproven skiljer sig en aning från de andra två ämnena genom att både delproven A och C erhåller något lägre värden.

Tabell 4. Sammanställning av hur data passar den hierarkiska modellen i enlighet med Figur 1 för nio nationella prov i NO-ämnen för årskurs 6.

Prov	CFI	TLI	RMSEA
Biologi 2013	0,97	0,97	0,028
Biologi 2014	0,98	0,98	0,024
Biologi 2015	0,98	0,98	0,024
Fysik 2013	0,96	0,96	0,028
Fysik 2014	0,94	0,93	0,035
Fysik 2015	0,94	0,94	0,040
Kemi 2013	0,96	0,95	0,036
Kemi 2014	0,98	0,98	0,023
Kemi 2015	0,92	0,92	0,075
Kemi 2015 (-C13)	0,98	0,97	0,024

Not: CFI = Comparative Fit Index; TLI = Tucker Lewis Index; RMSEA = Root Means-Square Error of Approximation

Modellen passar således data ännu lite bättre än enfaktormodellen som redovisas i Tabell 3. Fysikproven 2014 och 2015 ligger nära referensvärdena när det gäller CFI och TLI (Hu & Bentler, 1999) (se Tabell 4).

Kontroll av dimensionalitet för Rasch-modeller av proven

För att undersöka om uppgifterna i proven uppfyller kravet på endimensionalitet med hjälp av IRT, det vill säga att varje prov i huvudsak mäter en kompetens, används Rasch-modellering. Ett sätt att kontrollera om data passar modellen är att beräkna Weighted Mean Square (Bond & Fox, 2001). Rekommenderade värden, för att uppfylla kravet på att modellen passar data, är mellan 0,6 och 1,4 för uppgifter med öppna svar som är manuellt bedömda, och mellan 0,8 och 1,2 för flervalsfrågor. Uppgifterna, såväl slutna som öppna, i alla de nio proven uppfyller med några få undantag de skarpare rekommendationerna.

Vidare har principalkomponentanalys genomförts av Rasch-residualerna. Denna analys visar att det finns några mindre komponenter utöver huvudfaktorn i Raschmodellen. Dessa komponenter hänger samman med bland annat uppgiftsformat, det vill säga om uppgifterna innebär att eleverna behöver läsa och/eller skriva mycket eller om det rör sig om kortare uppgifter där eleverna tar ställning till en kortfattad fråga genom att sätta ett kryss. Det går också att urskilja en komponent som särskiljer delproven A och B från delproven C. Att särskilja delproven A från delproven B har inte visat sig vara möjligt. Ett skäl till detta kan vara att framförallt delproven A består av få uppgifter. Utöver detta går det också att, i vissa prov, urskilja några begreppsliga innehåll. Ett exempel från fysik är delprov C 2014 där uppgifter som handlade om elektricitet bildar en grupp i analysen, samtidigt som motsvarande innehåll i delprov C år 2015 inte urskiljer sig på samma sätt. Ett annat exempel är uppgift C13 i kemiprov 2015 som handlar om massans bevarande vid fasövergångar och som teoretiskt sett tillhör konstruktet, men uppgiften tycks ställa andra krav på eleverna än övriga uppgifter.

Konsekvenser av låg reliabilitet

Ett sätt att uppskatta reliabilitet är att beräkna Koefficient Alpha, som kan anta värden mellan noll och ett, där ett står för perfekt reliabilitet. Alpha-värdet utgör ett mått på intern konsekvens i provet. I Tabell 5 framgår Koefficient Alpha för de nio delproven A mellan 2013 och 2015.

Tabell 5. Reliabilitet för delprov A i de nio proven uttryckt som Koefficient Alpha

	2013	2014	2015
Biologi	0,62	0,68	0,53
Fysik	0,63	0,69	0,65
Kemi	0,63	0,65	0,63

Med utgångspunkt i reliabilitetsvärdena i Tabell 5 kan ytterligare ett sätt att förstå mätfelens inverkan beräknas, ett sätt som går ut på att uppskatta ett intervall inom vilket det sanna resultatet ligger. Med en elevs ”sanna” resultat menas det resultat som skulle erhållits om det inte fanns några felaktigheter i mätningen, och som därför avspeglar elevens egentliga kunskapsnivå. För ett resultat på fem poäng (av nio möjliga) på ett specifikt delprov A (Fysik 2013) med reliabilitet på 0,63 ligger det sanna resultatet med 95 %-ig säkerhet mellan 2,7 och 7,3 poäng. Det betyder alltså att mätfelet för en enskild elev som erhållit 5 poäng omfattar halva antalet möjliga poäng på delprovet. Se Tabell 6 för konfidensintervall för alla delprov A.

Tabell 6. Det 95 %-iga konfidensintervallet för ett delprovresultat mitt på skalan (5 av 9 poäng).

	2013	2014	2015
Biologi	2,3 – 7,7	2,8 – 7,3	2,5 – 7,5
Fysik	2,7 – 7,3	3,7 – 10,3	2,2 – 7,8
Kemi	2,5 – 7,4	2,8 – 7,2	2,6 – 7,3

Generellt går det att konstatera att det 95 %-iga konfidensintervallet omfattar i storleksordningen halva skalan för samtliga delprov A. Om till exempel gränser för olika provbetygsnivåer för detta delprov skulle sättas till 3 poäng för E, 5 poäng för C och 7 poäng för A skulle det 95%-iga konfidensintervallet omfatta alla betygssteg.

Sammanfattning av resultat

Resultaten från analyserna visar att det i huvudsak är en dimension av elevernas kunskaper som prövas i de nationella proven i NO-ämnena för årskurs 6. Både en enfaktormodell och en hierarkisk modell med en överordnad och tre underliggande faktorer passar data bra. Olika uppgiftsformat i de olika delproven kan vara en möjlig delförklaring till att den hierarkiska modellen ger något bättre modellpassning än enfaktormodellen.

Våra analyser visar också att en uppdelning av proven i delprov leder till att reliabiliteten för de minsta delproven blir för låg för att delprovsbetyg ska vara lämpligt att använda. Det innebär att det är problematiskt att införa villkor som tillmäter delproven en avgörande betydelse vid bestämning av provbetygen.

Diskussion

Vår första forskningsfråga handlade om huruvida det går att på ett tillförlitligt sätt urskilja de tre ”förmågorna” (”kommunicera”, ”undersöka” respektive ”använda begrepp”) i datamaterialet, eller om det är rimligare att betrakta resultaten som ett uttryck för en enda överordnad dimension. Analyserna ledde fram till två möjliga modeller för proven. Den ena är en enfaktormodell där elevernas svar antas bero av ett endimensionellt konstrukt. Den andra modellen är hierarkisk med en överordnad och tre underliggande faktorer, en för varje ”förmåga” som har prövats med varsitt delprov. Data stämmer väl överens med båda modellerna, men modelpassningen är något högre för den hierarkiska modellen. Proven kan därmed tyckas pröva de tre aspekter av naturvetenskaplig kunskap som kursplanerna lyfter fram och som är i linje med den internationella utvecklingen av kursplaner i de naturvetenskapliga ämnena (Kind, 2013; Orpwood, 2007; Stuckey m.fl., 2013). Emellertid är det tydligt att dessa tre aspekter korrelerar högt med varandra genom en överordnad faktor. Sammantaget gör detta att det inte har gått att på ett tillförlitligt sätt urskilja de tre ”förmågorna” i elevernas svar på proven, i betydelsen att de skulle vara enskilda entiteter oberoende av varandra. Detta trots att proven var konstruerade med avsikten att pröva elevernas kunskaper mot de specifika delar av kunskapskraven som relaterar till varje ”förmåga”. Istället har det visat sig rimligare att betrakta elevernas prestationer på proven som aspekter av en och samma övergripande faktor som kanske kan benämnas skolämneskunskap i biologi, fysik respektive kemi.

Dimensionaliteten i proven har också undersökts genom principalkomponentanalyser av Rasch-residualerna. Ett intressant resultat från dessa är att elevernas prestationer i delproven C håller sig inom en dimension och att elevernas prestationer i delproven A och B tillsammans håller sig inom en dimension. Det har inte varit möjligt att urskilja resultaten från delproven A från B på detta sätt. Analysen av Rasch-residualerna indikerar också att andra aspekter hos elevernas kunskaper har viss betydelse, vilket visar sig genom resultat från uppgifter med olika uppgiftsformat. Den övervägande

andelen uppgifter som innehåller relativt mycket text att läsa och som kräver förhållandevis mycket textproduktion från eleverna finns i delproven A och B. Därmed är det troligt att elevernas kunskaper när det gäller att läsa och skriva inverkar olika mycket på de olika delproven. Ett stöd för denna tolkning är att flickor, som anses läsa och skriva bättre än pojkar (Bolger & Kellaghan, 1990; Eriksson, 2005), i de nationella proven presterar bättre på uppgifter med öppna svarsformat och att det slår igenom när det gäller hur väl de presterar på de olika delproven (Bach m.fl., 2015, s. 59–61). En fråga som aktualiseras genom detta är i vilken utsträckning läs- och skrivförmåga skall anses vara viktiga aspekter av konstruktet och hur mycket de ska tillåtas påverka provbetygen. Förändringen i synen på kunskaper i skolämnena biologi, fysik och kemi skulle kunna motivera att konstruktet bör omfatta mer av det som traditionellt har betraktats som läs- och skrivförmåga. Det är dock inte helt givet att till exempel förmågan att ”kommunicera” innebär att en person måste vara duktig på att läsa och/eller skriva. Den fullständiga lydelsen för förmågan att ”kommunicera” är ju betydligt bredare än att bara handla om skriftlig kommunikation, och handlar om att använda kunskaper i naturvetenskap ”för att granska information, kommunicera och ta ställning…”, vilket inte nödvändigtvis måste ske skriftligt.

En av uppgifterna i kemiprovet från 2015 visade delvis unika egenskaper, uppgift C13. Den handlar om massans bevarande vid fasövergångar och behandlar därmed ett område som tidigare har visat sig vara rejält utmanande för elever i olika åldrar (Andersson, 1990; Driver, 1985; Galili & Bar, 1997). Vår hypotes är att det begrepp som prövas i uppgiften är så specifikt att det inte kan förstås med någon slags allmän kemikunskap, utan att det snarare kräver att eleven har träffat på, och kanske undervisats, om just detta. Trots att uppgiften inte riktigt passar in i modellen och uppvisar en avvikande dimensionalitet finns det goda argument för att den ska kunna vara med i ett prov då den avser att pröva elevernas kunskaper om en grundläggande princip i kemi. På ett generellt plan kan man konstatera att uppgifter som prövar olika delar av ett ämnesinnehåll med automatik leder till ett visst mått av flerdimensionalitet. Kanske kan det till och med vara så att alltför höga ambitioner att nå endimensionalitet vid provkonstruktion riskerar att leda till att speciella delar av ett ämnesinnehåll undviks, alltså det man brukar kalla underrepresentation av konstruktet?

Analyserna av reliabilitet visar att osäkerheten i provbetygen skulle bli betydande ifall de avgjordes på ett icke-kompensatoriskt sätt av delprovsbetyg. Validiteten hos ett sådant provbetyg måste kraftigt ifrågasättas. Det är till exempel inte särskilt troligt att en elev skulle erhållit samma provbetyg om eleven hade genomfört en annan av de teoretiskt sett oändligt många varianterna av provet. En tanke som har framförts är att man skulle kunna använda delprovsbetyg vid kommunikation av elevers kunskapsprofil, men inte heller det kan betraktas som en valid användning av delprovsbetyg med tanke på den låga reliabiliteten. En sådan användning skulle innebära att provens omfattning och därmed den tid som behövs för att genomföra och bedöma proven behöver utökas relativt kraftigt.

Provbetygen i NO-proven bör, både av reliabilitetsskäl och på grund av hög korrelation mellan faktorerna, sättas enligt en kompensatorisk princip där goda prestationer inom vissa delar av provet uppväger mindre goda prestationer inom andra delar av provet. Ett sådant förfarande står delvis i kontrast med det svenska betygssystemet. Skolverket beskriver detta som ett system där kunskapskravet i sin helhet måste vara uppfyllt för betyg motsvarande en bestämd nivå; E, C eller A. För de mellanliggande nivåerna D och B ska överliggande nivå till övervägande delen vara uppfyllt (Skolverket, 2016a).

Våra resultat visar att de nationella proven i NO-ämnena för årskurs 6 inte lyckas generera tillförlitlig information om enskilda elevers kunskaper i de tre olika ”förmågorna”. Är det då rimligt att förvänta sig att lärarna ska kunna få fram tillförlitlig information om sina elevers kunskaper i förhållande till var och en av ”förmågorna”? Eller, för att ställa problemen med att vid bedömning dela upp kunskap i bitar på sin spets; Är det rimligt att förvänta sig att lärarna ska kunna få fram tillförlitlig information till varje formulering i kunskapskraven, med tanke på att de har minst 12 separata kravformuleringar i varje NO-ämne att ta i beaktande? Ett icke-kompensatoriskt sätt att arbeta med betygssättning

kräver att läraren måste sträva efter att undersöka vilken av alla kravformuleringar eleven uppfyller, så att säga, sämst. Denna nivå definierar då det betyg som ska sättas, oavsett hur väl andra delar av kunskapskravet är uppfyllda. Om en lärare ska göra detta rättssäkert krävs många indikationer för varje kravformulering. Undervisningen i exempelvis kemi omfattar i storleksordningen 20 timmar under årskurs sex. Om alla separata kravformuleringar ska hanteras inom ramen för de 20 timmarna med flera oberoende bedömningar lär det inte bli mycket tid över till undervisning.

Denna studie har visat att en uppdelning av de nationella proven i delprov leder till validitetsproblem på grund av brist på reliabel information om enskilda elevers prestation på varje del. En lärare hamnar obönhörligen i samma typ av validitetsproblem om hon eller han inte har tillräckligt reliabel information om varje separat kravformulering i kunskapskraven. Eleverna å sin sida riskerar att bli bedömda på varje liten insats för att läraren ska kunna motivera sin betygsättning om underlag efterfrågas av elev, föräldrar eller rektor. Resultaten i den här studien visar att det är möjligt att förstå kunskaper i de naturvenskapliga ämnena som varsitt samlat konstrukt, kunskaper i biologi, kunskaper i fysik respektive kunskaper i kemi. Om betygssystemet tillät lärare att förstå kunskaper i varje skolämne som en helhet, istället för som lika många delar som kravformuleringar, skulle förutsättningarna för lärares och elevers arbete förbättras avsevärt. Att betygssystem påverkar skolans arbete är i det närmaste självklart. Det har också varit känt sedan länge att ett betygssystem kan stå i konflikt med skolans uppdrag att bedriva utbildning. Frits Wigforss varnade för detta redan 1942 när han beskrev ett målrelaterat betygssystem: ”*Det är en svaghet i den diskuterade metoden, att ju effektivare den blir ur betygssynpunkt, dess större blir risken för skadlig inverkan på skolarbetet…*” (Regeringen, 1942, s. 53).

Tack

Avslutningsvis önskar vi framföra ett stort tack till de två granskarna för utförliga, insiktsfulla och hjälpsamma kommentarer.

Referenser

Andersson, B. (1990). Pupils’ conceptions of matter and its transformation (age 12-16). *Studies in Science Education*, (18), 53–58. https://doi.org/10.1080%2F03057269008559981

Bach, F., Frändberg, B., Hagman, M., West, E., & Zetterqvist, A. (2015). De nationella proven i NO åk 6 Skillnader i resultat mellan olika grupper. *Educare*, *2015*(1). Hämtad från https://www.mah.se/upload/FAKULTETER/LS/Publikationer/EDUCARE/Educare%202015.2.pdf

Bolger, N., & Kellaghan, T. (1990). Method of Measurement and Gender Differences i Scholastic Achievemnt. *Journal of Educational Measurment*, *27*(2), 165–174. https://doi.org/10.1111/j.1745-3984.1990.tb00740.x

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch Model. Fundamental Measurement in the Human Science*. Mahwah, NJ: Lawrence Erlbaum Associates. https://doi.org/10.1111/j.1745-3984.2003.tb01103.x

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (Second edition.). London: The Guilford Press.

Chou, Y.-T., & Wang, W.-C. (2010). Checking Dimensionality in Item Response Models With Principal Component Analysis on Standardized Residuals. *Educational and Psychological Measurement*, *70*(5), 717–731. https://doi.org/10.1177/0013164410379322

Crocker, L., & Algina, J. (2008). *Introduction to Classical and Modern Test Theory*. Cengage Learning.

Driver, R. (1985). Beyond appearance: the conservation of matter. I R. Driver, E. Guesne, & A. Tiberghien (Red.), *Children’s Ideas in Science* (s. 145–169). Milton Keynes: Open University Press.

Eriksson, N. (2005). *Prestationsskillnader mellan flickor och pojkar i NO: En studie av uppgiftsformatets betydelse i TIMSS 2003* (BVM No. 15) (s. 25). Umeå: Enheten för beteendevetenskapliga mätningar. Hämtad från http://www.jus.umu.se/digitalAssets/59/59497_bvm-nr-15.pdf

Galili, I., & Bar, V. (1997). Children's operational knowledge about weight. *International Journal of Science Education*, *19*(3), 317–340. https://doi.org/10.1080/0950069970190305

Gustafsson, J.-E. (2008). Effects of International Comparative Studies on Educational Quality on the Quality of Educational Research. *European Educational Research Journal*, *7*(1), 1–17. https://doi.org/10.2304/eerj.2008.7.1.1

Gustafsson, J.-E. (2009). Strukturella ekvationsmodeller. I G. Djurfeldt & M. Barmark (Red.), *Statistisk verktygslåda 2: multivariat analys* (1. uppl.). Stockholm: Studentlitteratur.

Gustafsson, J.-E., Cliffordsson, C., & Erickson, G. (2014). *Likvärdig kunskapsbedömning i och av den svenska skolan – problem och möjligheter* (s. 120). SNS - Studieförbundet Näringsliv och Samhälle. Hämtad från https://www.sns.se/aktuellt/likvardig-kunskapsbedomning-i-och-av-den-svenska-skolan-problem-och-mojligheter/

Heine, J.-H. (2017). *Pairwise* (Version 0.4.3-2). Hämtad från https://cran.r-project.org/package=pairwise

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Kind, P. M. (2013). Conceptualizing the Science Curriculum: 40 Years of Developing Assessment Frameworks in Three Large-Scale Assessments. *Science Education*, *97*(5), 671–694. https://doi.org/10.1002/sce.21070

Muthén, B., & Muthén, L. (2016). *Mplus (Version 7.4)*. Hämtad från http://www.statmodel.com

Orpwood, G. (2001). The Role of Assessment in Science Curriculum Reform. *Assessment in Education: Principles, Policy & Practice*, *8*(2), 135–151. https://doi.org/10.1080/09695940125120

Orpwood, G. (2007). Assessing Scientific Literacy: Threats and Opportunities. I C. Linder, L. Östman, & P.-O. Wickman (Red.), *Promoting Scientific Literacy: Science Education Research in Transaction* (s. 120–129). Hämtad från https://eprints.qut.edu.au/12883/1/Linnaeus_Tercentenary_Symposium.pdf

Pedhazur, E., & Pedhazur Schmelkin, L. (1991). *Measurement, Design, and Analysis. An integrated Approach*. New York: Psychology Press.

R Development Core Team. (2008). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. Hämtad från http://www.R-project.org

Regeringen. (1942). *Betänkande med utredning och förslag angående betygssättningen i Folkskolan*. Hämtad från http://weburn.kb.se/metadata/980/SOU_8219980.htm

Regeringen. (2017, september 14). *Nationella prov – rättvisa, likvärdiga, digitala* [Text]. Hämtad 01 mars 2018, från http://www.regeringen.se/rattsdokument/proposition/2017/09/prop.-20171814/

SFS 2011:185. *Förordning om ändring i skolförordningen*. *S*tockholm: Utbildningsdepartementet. Hämtad från https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/skolforordning-2011185_sfs-2011-185

Sjaastad, J. (2014). Enhancing measurement in science education research through Rasch analysis: Rationale and properties. *Nordic Studies in Science Education*, *10*(2), 212–230. https://doi.org/10.5617/nordina.662

Skolverket. (2016a). *Betygsskalan och betygen B och D*. Stockholm: Skolverket. Hämtad från https://www.skolverket.se/publikationer?id=2953

Skolverket. (2016b). *Utvärdering av den nya betygsskalan samt kunskapskravens utformning*. Hämtad från https://www.skolverket.se/publikationer?id=3652

Skolverket. (2018). *Läroplan för grundskolan, förskoleklassen och fritidshemmet 2011*. Stockholm: Skolverket. Hämtad från https://www.skolverket.se/publikationer?id=3813

Stuckey, M., Hofstein, A., Mamlok-Naaman, R., & Eilks, I. (2013). The meaning of 'relevance' in science education and its implications for the science curriculum. *Studies in Science Education*, *49*(1), 1–34. https://doi.org/10.1080/03057267.2013.802463

Yates, L., & Collins, C. (2010). The Absence of Knowledge in Australian Curriculum Reforms. *European Journal of Education*, *45*(1), 89–102. https://doi.org/10.1111/j.1465-3435.2009.01417.x