

ÖKAR RELIABILITETEN VID DIFFERENTIELL POÄNGSÄTTNING?

En studie av delproven NOG och LÄS i högskoleprovet

**Anders Lexelius
Stig Eriksson**

Pm nr 170, 2001



ISSN 1100-696X
ISRN UM-PED-PM--170--SE

Högskoleprovet består av fem delprov. Varje individ som deltar i högskoleprovet erhåller en provpoäng som utgör summan av antalet rätt besvarade uppgifter i provet. I två av dessa delprov NOG och LÄS är vissa svarsförslag mer felaktiga än andra, dvs olika felsvar indikerar olika kunskapsnivåer. I denna typ av delprov borde en tillämpning av en differentiell poängsättning kunna tillvarata den informationen som finns i svaren på ett mer effektivt sätt än dikotom poängsättning. Detta låter sig dock inte enkelt göras. Med differentiell poängsättning menas att man ger poäng för flera än ett svarsförslag efter en fallande skala där det svarsförslag som är mest rätt ger högst poäng. Det är inte ovanligt att provpoängen för en individ varierar mellan olika provtillfällen. För varje provtagare kan provpoängen delas upp i en "sann poäng" och en "felpoäng".

$$\text{provpoäng} = \text{sann poäng} + \text{felpoäng}$$

Orsaken till variationen i provpoängen kan vara dagsform, gissningsförmåga etc. Den *sanna* poängen är däremot, under likartade betingelser, alltid densamma för individen. Utbildning och kunskap är dock två faktorer som påverkar den sanna poängen i positiv riktning. En relevant fråga man kan ställa sig i detta sammanhang är hur man kan göra uttalanden om individens sanna poäng när den är okänd. Skillnaden mellan provpoängen och den sanna poängen utgörs av felpoängen. Hur stort inflytande har då felpoängen på provpoängen? Om man kan besvara den frågan så kan man uttala sig om provets reliabilitet. Klart är dock att ju mindre felpoäng desto högre reliabilitet har provet.

Nu är det inte bara provpoängen som kan delas upp i sann poäng och felpoäng, utan detsamma gäller även för variansen under vissa förutsättningar¹.

$$\text{variansen i provpoäng} = \text{variansen i sann poäng} + \text{variansen i felpoäng}$$

Man ökar reliabiliteten i ett prov genom att andelen varians i sann poäng ökar proportionellt mer än provets totala varians.

$$\text{reliabiliteten} = (\text{variansen i sann poäng}) / (\text{variansen i provpoäng})$$

¹ 1. Medelvärdet för felpoängen i provet = 0. 2. Korrelationen mellan felpoäng och sann poäng = 0

I provsammanhang är det vanligt att provets uppgifter är relativt högt korrelerade och att varje uppgift dessutom bidrar med lika stor del av totalpoängen. Konkret innebär det att provet är endimensionellt, dvs. att uppgifterna i provet i huvudsak mäter en och samma förmåga. Detta är även något som borgar för att provet har en hög reliabilitet. Flera dimensioner i ett prov betyder som regel en sänkning av reliabiliteten pga. att prestationen varierar på de uppgifter som representerar provets olika bakomliggande dimensioner.

För att prov ska vara användbara i urvalssammanhang krävs att de är reliabla och valida, dvs. provpoängen ska vara tillförlitlig och stabil. Två centrala begrepp inom såväl den klassiska som den moderna test-teorin är *reliabilitet* och *validitet*. Kortfattat kan man sammanfatta reliabiliteten som ett mått på *hur stabila* mätvärdena i ett test är. Om man upprepar ett och samma test flera gånger bör man erhålla samma mätresultat, under förutsättning att betingelserna är likartade, för att ett test ska ha hög reliabilitet. Naturligtvis begränsas förfarandet av att man i praktiken inte kan ge samma test till en och samma grupp utan att ett tidigare provtagande påverkar det efterföljande. Validiteten svarar på frågan *vad* testet mäter och är en process där man försöker få stöd för den *tolkning* som ska göras utifrån testpoängen.

I rättningssammanhang är det vanligt att belöna varje rätt besvarad uppgift med 1 poäng medan felaktigt besvarade och överhoppade uppgifter inte får någon poäng.

Ett sätt att höja reliabiliteten hos ett prov är att öka andelen sann varians på bekostnad av felvariansen. Rent praktiskt skulle detta kunna åstadkommas på två olika sätt: 1) genom att *öka antalet uppgifter i provet* eller 2) genom att tillämpa en *differentiell poängsättning*, där kunskap av mer eller mindre rätta svarsförslag belönas i ett (på förhand/efterhand) fastslaget poängsystem. För att en höjning av reliabiliteten ska bli möjlig enligt matematikens lagar, måste också "rätt" personer belönas i den differentiella poängsättningen. Sålunda får inte de sämre individerna belönas på de svåraste uppgifterna. Det ger i så fall en sänkning av reliabiliteten eftersom spridningen i totalpoäng minskar. I motsats till dikotom poängsättning påverkas reliabiliteten av hur stort avståndet är mellan de olika poänggränserna vid differentiell poängsättning. Vid differentiell poängsättning kan poängvikterna antingen bestämmas i förväg, dvs. *a priori*, eller i efterhand på *empi-*

risk väg. Då vikterna bestäms i efterhand gör man skillnad mellan vikter som utnyttjar ett *internt* kriterium (provpoäng) och sådana som utnyttjar ett *externt* kriterium (betyg, poäng på parallellt prov).

Tidigare forskning

Coombs, Milholland & Womer (1956) har funnit visst stöd för att partiell kunskap existerar och att reliabiliteten i flervalssprov kan höjas givet denna kunskap. Höjningen motsvarade i genomsnitt en förlängning av ett konventionellt rättat prov, enligt Spearman-Browns formel, med en faktor 1,2.

Davis & Fifer (1959) erhöll en kraftig reliabilitetshöjning vid differentiell poängsättning med empiriska vikter, jämfört med konventionell rättning. Davis & Fifer visade att flervalssprovets reliabilitet maximerades då vikterna stod i direkt proportion till medelvärdet för de individer som valt respektive svarsförslag.

Aiken (1967) har matematiskt härlett variansuttrycken under olika poängsättningstekniker. Han menar att den reliabilitetshöjning man erhåller vid differentiell poängsättning står i ett direkt förhållande till den variansökning man erhåller vid hänsynstagande till partiell kunskap. Aiken fann att testpoängvariansen vid differentiell poängsättning varierar direkt med:

- ❑ Antalet uppgifter i testet
- ❑ Antalet svarsförslag per uppgift
- ❑ Kvadraten av de olika responsvikterna
- ❑ Proportionen individer som väljer ett givet svarsförslag

Huruvida reliabiliteten ökar vid differentiell poängsättning framgår dock inte direkt av Aikens resonemang. Vad som krävs för att reliabiliteten ska öka vid differentiell poängsättning är att summan av alla kovarianstermer ska öka proportionellt mer än summan av uppgiftsvarianserna.

Wedman (1969) erhöll både en ökning och en minskning av reliabiliteten när han undersökte effekten av differentiell poängsättning på ett prov som avsåg kunskaper i samhällsorientering på högstadiets avgångsklass i grundskolan. Ökningen från 0,515 till 0,722 erhöles via empiriska vikter med biseriala korrelationskoefficienter. Vid en kors-

validering på ett nytt stickprov och som utfördes för att pröva de tidigare vikterna erhöles en sänkning av reliabiliteten från 0,584 till 0,556. Wedman (1970) har också undersökt reliabiliteten med hjälp av två synonymordprov bestående av 40 uppgifter vardera i årskurserna 5 och 6 i grundskolan. Provens viktning kom att grunda sig på både subjektiva bedömningar och empiriska resultat. Vid viktning a priori erhöles vissa positiva effekter på reliabiliteten medan ett empiriskt viktningförfarande inte påverkade reliabiliteten.

Även Backman & Wedman (1971) fann i en reliabilitetsstudie av två synonymordprov att resultaten pekar åt samma håll som de två tidigare undersökningarna (Wedman 1969 och 1970). Återigen kunde alltså en reliabilitetsminskning noteras.

Vid en enligt Wedman, mycket utförlig litteratursammanställning, fann Stanley & Wang (1968) och Wang & Stanley (1970) att forskningsresultaten inte är entydiga vad beträffar reliabilitetshöjningar vid differentiell poängsättning.

I ett flertal undersökningar där variabel viktningsteknik har använts och poängsättningen därmed skett på olika vis har inte automatiskt några reliabilitetshöjningar observerats jämfört med konventionell rättning (Rippey, 1968, 1970; Ebel 1968, Hambleton, Roberts, Traub & Koehler 1971).

Hendrickson (1971) erhöles i en mycket omfattande studie genomgående mycket positiva resultat vid poängsättning baserad på empiriska vikter. I genomsnitt motsvarade den reliabilitetshöjning hon erhöles med empiriska vikter en förlängning av de konventionellt poängsatta proven med en faktor 1,5.

I en studie där delprovet NOG utnyttjades som referensprov och där det övergripande syftet var att via en empirisk prövning erhålla mättekniska indikatorer på ett nykonstruerat numeriskt prov erhöles en reliabilitetshöjning på NOG-provet från 0,43 till 0,59 med differentiell poängsättning (Henriksson, Lexelius & Hamrén 1985). NOG-provet viktades a priori med en poängsättningsprincip som utarbetats av Henriksson (1981). Det bör noteras att reliabiliteten 0,43 är mycket låg för att vara ett NOG-prov och att uppgifterna i ett prov med så låg relia-

bilitet inte uppfyller de krav som ställs på ett reguljärt delprov i högskoleprovet.

Efter att ha tagit del av flertalet undersökningar angående differentiell poängsättning kan man konstatera, dels att nya studier förekommer sparsamt, dels att tidigare forskning, gett motstridiga resultat. Det senare innebär att man via en differentiell poängsättningsteknik inte automatiskt bidrager till reliabilitetshöjningar. Värt att notera är även att de prov och försökspersoner som deltagit i försöken med olika viktning förfaranden inte varit representativa vare sig för högskoleprovet eller för högskoleprovets deltagare. Flertalet av undersökningarna har varit av experimentell karaktär där man med olika tekniker räknat fram empiriska vikter för att erhålla reliabilitetshöjningar.

Syfte

Syftet med föreliggande studie är att undersöka hur differentiell poängsättning påverkar provpoängens reliabilitet i delproven NOG och LÄS i högskoleprovet. Ett annat syfte är att undersöka hur många uppgifter respektive prov måste förlängas med, för att man ska erhålla samma reliabilitet som vid differentiell poängsättning.

Metod

Med utgångspunkt i vårt syfte har vi, i föreliggande studie, valt att undersöka hur reliabiliteten påverkas av differentiell poängsättning i delproven NOG och LÄS, som ingår i högskoleprovet. Reliabilitetsstudien på NOG-provet genomfördes i samband med de två reguljära högskoleproven 1999. Motsvarande studie på LÄS-provet genomfördes hösten samma år. Med hjälp av Spearman-Browns förlängningsformel undersöktes hur reliabiliteten skulle förändras om antalet uppgifter i både NOG- och LÄS-provet utökades.

NOG-provet

NOG-provet är ett kvantitativt prov som avser att mäta numerisk-logisk förmåga. Provet består av 22 uppgifter av skiftande karaktär. Den uppgift som föreläggs provdeltagarna är att bedöma om tillräcklig information föreligger för att ett visst givet problem ska kunna lösas entydigt. Uppgiftsformatet i NOG-provet är fast, vilket innebär att svarsförslagen är identiska för alla uppgifter i provet.

Provet förutsätter kunskap om vissa grundläggande kvantitativa begrepp och enheter som definieras av kursplanen i matematik upp till och med kurs A för gymnasieskolan.

Nedan visas ett exempel på en NOG-uppgift för att illustrera svarsförslagets innebörd.

Ex

Sammanlagt deltog 792 kvinnor och män i en kongress. **Hur stor andel av deltagarna var kvinnor?**

(1) I kongressen deltog 396 färre män än kvinnor.

(2) I kongressen deltog 3 gånger så många kvinnor som män.

A i (1) men ej i (2)

B i (2) men ej i (1)

C i (1) tillsammans med (2)

D i (1) och (2) var för sig

E i *ej* genom de båda påståendena

Tabell 1. Svartsförslagens innebörd i nogprovet

Svartsförslag	Innebörd	Betyder
A	i (1) men ej i (2)	Den information som ges i (1) är tillräcklig. Enbart informationen i (2) räcker inte till.
B	i (2) men ej i (1)	Den information som ges i (2) är tillräcklig. Enbart informationen i (1) räcker inte till.
C	i (1) tillsammans med (2)	För att få tillräcklig information <i>måste</i> man använda både påstående (1) och (2). Enbart (1) eller enbart (2) ger ej tillräcklig information.
D	i (1) och (2) var för sig	Antingen (1) eller (2) kan användas, eftersom båda var för sig innehåller tillräckligt mycket information.
E	<i>ej</i> genom de båda påståendena	Inte ens genom att nyttja (1) och (2) kan man få tillräcklig information.

Varje korrekt besvarad uppgift i högskoleprovet, oavsett delprov, belönas med 1 poäng medan alla felaktiga svar och ej besvarade uppgifter ger 0 poäng. Genom att addera poängen för alla uppgifter erhålls en totalpoäng för hela provet.

Vid differentiell poängsättning kan man ge poäng för samtliga svartsförslag. Poängsättningen antas då återspegla olika grader av kunskap, dvs olika grader av rätt svar. Det svartsförslag som är mest rätt ger högst poäng och följaktligen ger det svartsförslag som är minst rätt lägst poäng.

Den modell som används vid poängsättning av NOG-provet har utarbetats av Henriksson (1981). NOG-uppgifter vars korrekta svartsförslag är A, B eller D erhöll maximalt 2 poäng, medan uppgifter vars korrekta svartsförslag är C eller E erhöll maximalt 3 poäng. Orsaken till denna poängskillnad är att A-, B- och D-alternativen tar ställning till två delproblem, medan C- och E-svaren tar ställning till tre delproblem. Uppgifterna är poängsatta så att varje korrekt delproblem ger 1 poäng. Antalet beslut där vikterna (poängen) är bestämda *a priori* framgår av tabell 2.

Tabell 2. Poäng för varje angivet svar i förhållande till korrekt alternativ med differentiell poängsättning i NOG-provet.

Korrekt alternativ	Angivet svar	Beslut (fet stil = rätt beslut)	Poäng
A (1) ja (2) nej	A	(1) ja (2) nej	2
	B	(1) nej (2) ja	0
	C	(1) nej (2) nej (1+2) ja	1
	D	(1) ja (2) ja	1
	E	(1) nej (2) nej (1+2) nej	1
B (1) nej (2) ja	A	(1) ja (2) nej	0
	B	(1) nej (2) ja	2
	C	(1) nej (2) nej (1+2) ja	1
	D	(1) ja (2) ja	1
	E	(1) nej (2) nej (1+2) nej	1
C (1) nej (2) nej (1+2) ja	A	(1) ja (2) nej	1
	B	(1) nej (2) ja	1
	C	(1) nej (2) nej (1+2) ja	3
	D	(1) ja (2) ja	0
	E	(1) nej (2) nej (1+2) nej	2
D (1) ja (2) ja	A	(1) ja (2) nej	1
	B	(1) nej (2) ja	1
	C	(1) nej (2) nej (1+2) ja	0
	D	(1) ja (2) ja	2
	E	(1) nej (2) nej (1+2) nej	0
E (1) nej (2) nej (1+2) nej	A	(1) ja (2) nej	1
	B	(1) nej (2) ja	1
	C	(1) nej (2) nej (1+2) ja	2
	D	(1) ja (2) ja	0
	E	(1) nej (2) nej (1+2) nej	3

Resultat NOG-provet

Studie I

Vid genomförandet av högskoleprovet våren 1999 deltog drygt 68 000 prövande. Resultatet på NOG-provet var lägre än normalt. Medelpoängen på de 22 uppgifterna var endast 9,92 med en spridning på 3,98. Vilket framgår av tabell 3 tabell vilken även inkluderar medelvärde och standardavvikelse för samma prov med differentiell poängsättning.

Tabell 3. Medelvärde och spridning för NOG-provet med dikotom- och differentiell rättning, totalt och uppdelat på kön samt antal prövande.

	Dikotom rättning		Differentiell rättning		N
	M	S	M	S	
Totalt	9,93	3,98	37,82	6,95	68476
Män	11,14	4,10	39,88	6,86	31498
Kvinnor	8,89	3,56	36,07	6,52	36978

Av tabell 3 framgår att männen har lyckats bättre på NOG-provet både vid konventionell- och differentiell poängsättning. Männen har även större spridning i poängen oavsett poängsättningsystem än kvinnorna. Noterbart är att kvinnorna genom differentiell poängsättning ökat sin poäng med faktor 4,07 medan motsvarande ökning för männen är 3,81.

Korrelationen mellan medelvärdena för konventionell- och differentiell poängsättning är 0,94 vilket indikerar att rangordningen av individer i de båda rättningssystemen är tämligen likartad.

Tabell 4. Reliabilitet, totalt och för män respektive kvinnor, enligt Cronbachs Alpha med dikotom och differentiell rättning samt antalet individer (N).

	Dikotom rättning			Differentiell rättning		
	Totalt	Män	Kvinnor	Totalt	Män	Kvinnor
Reliabilitet	0,707	0,727	0,634	0,732	0,744	0,682
N	68476	31498	36978	68476	31498	36978

Av tabell 4 framgår att man får en *måttlig* ökning av reliabiliteten vid differentiell poängsättning jämfört med konventionell. Tillskottet vid differentiell poängsättning för totalgruppen är i denna undersökning 0,025 enheter. Man kan även notera att kvinnorna reducerar sin felvarians mer än männen vid ett differentiellt poängsättningsförfarande. Ökningen i reliabilitet är 0,048 för kvinnorna och 0,017 för männen. Men fortfarande har kvinnornas provpoäng en högre felvarians än männens, dvs. männen har genomgående högre reliabilitet på provet än kvinnorna oavsett hur provet rättas.

Om man istället för ett differentiellt poängsättningsförfarande utökade antalet uppgifter i provet, så skulle det behövas 25 uppgifter i provet,

mot nuvarande 22, för att erhålla samma reliabilitetsökning som man får vid differentiell poängsättning.

En orsak till att reliabiliteten inte ökar mer vid differentiell poängsättning torde dels vara att varje uppgift i provet redan är granskad av experter samt även att uppgifterna är valda bland redan prövade uppgifter med bra mättekniska värden. Variationen i felpoäng reduceras endast måttligt vid differentiell poängsättning på prov som redan har granskats av experter och som dessutom är sammansatta för ett visst syfte. För att erhålla samma reliabilitet av NOG-provet som vid differentiell poängsättning måste provet, enligt Spearman-Browns förlängningsformel, utökas med ytterligare 3 uppgifter. Det torde alltså vara en mer framkomlig väg att utöka antalet uppgifter i provet än att inleda en diskussion med de prövande angående differentiell poängsättning, i synnerhet om man ska använda sig av empirisk viktning.

Studie II

Hösten 1999 upprepades försöket med differentiell poängsättning på NOG-provet och studien kom att omfatta 1843 prövande i Västerbottens län, 845 män och 998 kvinnor. I tabell 5 redovisas resultatet av den studien.

Tabell 5. Reliabilitet, totalt och för män respektive kvinnor, enligt Cronbachs Alpha med dikotom och differentiell rättning samt antalet individer (N).

	Dikotom rättning			Differentiell rättning		
	Totalt	Män	Kvinnor	Totalt	Män	Kvinnor
Reliabilitet	0,813	0,812	0,776	0,798	0,792	0,770
N	1843	845	998	1843	845	998

Som tabell 5 visar är reliabiliteten för den totala gruppen redan initialt relativt hög, drygt 0,81. Vid differentiell poängsättning, enligt tidigare modell, erhålls inga reliabilitetshöjningar varken för totalgruppen eller för män respektive kvinnor. Man kan istället notera sänkningar av reliabiliteten. Förklaringen till detta är förmodligen provets redan höga reliabilitet vid dikotom rättning och att man vid differentiell poängsättning inte förklarar mer av felvariansen. Istället har felvariansen relativt ökat mer än den observerade variansen och detta leder automatiskt till en sänkning av reliabiliteten.

LÄS-provet

LÄS-provet prövar förmågan att tillgodogöra sig information i svensk text. Provet består av fem texter med varierande innehåll hämtade från olika områden. Till var och en av de fem texterna finns fyra frågor som prövar läsförståelsen i vid mening. Till varje fråga finns det fyra svarsförslag varav ett är korrekt. Endast den information som finns i texten ska användas för att besvara frågorna.

Vid differentiell poängsättning av LÄS-provet har *empirisk* viktning tillämpats, till skillnad från den a priori viktning som tillämpades i NOG-provet. Vikterna för varje svarsförslag (A, B, C och D) bestämdes utifrån den medelpoäng de svarande på respektive svarsförslag hade på hela LÄS-provet.

Det korrekta svarsförslaget tilldelades således 3 poäng. Felaktigt svarsförslag med det högsta medelvärdet gavs 2 poäng, det näst högsta 1 poäng och det svarsförslag med lägst medelvärde gavs 0 poäng. 0 poäng tilldelades även för obesvarad uppgift.

Resultat LÄS-provet

Som försöksprov användes det reguljära LÄS-provet från hösten 1999, med data enligt tabell 6.

Tabell 6. Medelvärde och spridning för LÄS-provet med dikotom- och differentiell rättning, totalt och uppdelat på kön samt antal prövande.

	Dikotom rättning		Differentiell rättning		N
	M	S	M	S	
Totalt	11,20	3,77	44,43	7,77	49693
Män	11,39	3,79	44,85	7,70	22153
Kvinnor	11,04	3,75	44,10	7,81	27540

Av tabell 6 framgår att männen i genomsnitt har en något högre poäng på LÄS-provet än kvinnorna både vid konventionell- och differentiell poängsättning. Männen har något högre spridning i poäng med ett dikotomt rättningsförfarande än kvinnorna. Med ett differentiellt rättningsförfarande har dock kvinnorna högre spridning än männen.

Korrelationen mellan medelvärdena för konventionell- och differentiell poängsättning är för Läs-provet 0,95 vilket indikerar att rangordningen av individer i de båda rättningsystemen är tämligen likartad.

Tabell 7. Reliabilitet, totalt och för män respektive kvinnor, enligt Cronbachs Alpha med dikotom och differentiell rättning samt antalet individer (N).

	Dikotom rättning			Differentiell rättning		
	Totalt	Män	Kvinnor	Totalt	Män	Kvinnor
Reliabilitet	0,708	0,714	0,703	0,707	0,709	0,705
N	49693	22153	27540	49693	22153	27540

Med ovan beskrivna empiriska viktning erhöles reliabiliteten 0,707 vid differentiell rättning mot 0,708 vid dikotom rättning, dvs knappast någon förändring över huvud taget, och definitivt ingen förbättring. Provets reliabilitet, för män respektive kvinnor, påverkas inte heller av vilket rättningsförfarande man väljer.

Anledningen till detta torde vara att den använda empiriska viktningen inte ändrar rangordningen mellan de prövande. Man kunde ha tänkt sig att istället på förhand värderat och viktat svarsförslagen utifrån innehållsmässig/kognitiv svårighetsgrad. Men för LÄS-provet ter sig detta närmast omöjligt. Det skulle ställa orimliga krav på granskarna och granskningsprocessen. Dessutom skulle det bli svårt att inför provtagarna försvara en sådan viktning, särskilt om den inte stämmer med de uppgiftsresultat man får efter provets genomförande.

Om man önskar att höja reliabiliteten hos LÄS-provet är sannolikt bästa lösningen – förmodligen även för NOG – att öka antalet uppgifter i provet. Detta är säkert genomförbart för NOG, utan ändringar i provformat och provtid. 2-4 uppgifter extra skulle nog fungera för de flesta provtagare. LÄS är å andra sidan redan alltför tidsberoende, och här skulle man i stället bli tvungen att vidta andra åtgärder som tex. att minska antalet långa texter och lägga till fler korta. Kanske även finna nya typer av uppgifter.

Sammanfattning

Vid differentiell poängsättning av delproven NOG och LÄS, där vikterna bestämdes *a priori* respektive *empiriskt*, erhöles endast måttlig eller ingen höjning av reliabiliteten jämfört med dikotom poängsättning. Resultatet är på intet sätt förvånande då varje uppgift i NOG respektive LÄS valts ut med största omsorg. Uppgifterna är dessutom granskade av experter som är väl förtrogna med såväl provens innehåll som bakomliggande provteori. Tidigare forskning ger inte heller något entydigt stöd för ökad reliabilitet vid differentiell poängsättning då resultaten går i olika riktningar. En mer framkomlig väg för att höja reliabiliteten i NOG- och LÄS-provet är att öka antalet uppgifter i respektive prov. Det kan göras utan vidare åtgärd inom nuvarande provtid för NOG-provet. För att kunna behålla nuvarande provtid för LÄS-provet måste man vidtaga åtgärder som att minska antalet långa texter till förmån för flera korta om man ska öka antalet uppgifter i provet.

Referenser

- Aiken, L.R., Jr. (1967). Effect on test score variance of differential weightings of item responses. *Psychological Reports*, 21, 585-590.
- Backman, J. & Wedman, I. (1971). *Differentiell poängsättning av flervalfrågor med viktsystem grundade på prestations- och perceptionsmätningar* (Pedagogiska rapporter, nr 21). Umeå universitet.
- Coombs, C.H., & Milholland, J.E. & Womer, F.B. (1956). The assessment of partial knowledge. *Educational and Psychological Measurement*, 16, 13-37.
- Davis, F.B., & Fifer, G. (1959). The effect on test reliability and validity on scoring aptitude and achievement test with weights for every choice. *Educational and Psychological Measurement*, 19, 159-170.
- Ebel, R.L. (1968). Valid confidence testing-demonstration kit. Rippey, R.M. Probabilistic Testing. *Journal of Educational Measurement*, 5, 353-354.
- Hambleton, R.K., Roberts, D.M. & Traub, R.E. A. (1970). Comparison of the reliability and validity of two methods for assessing partial knowledge on multiple-choice test. *Journal of Educational Measurement*, 7, 75-82.
- Hendrickson, G.F. (1971a). *An assessment of the effect of differentially weighting options of a multiple-choice objective test using a Guttman weighting scheme* (Working Paper, No 6). The John Hopkins University: Center for Social Organization of Schools.
- Hendrickson, G.F. (1971b). *The effect of differential option weighting on multiple-choice test* (Technical Report, No 93). The John Hopkins University.
- Hendrickson, G.F. (1971). The effect of differential option weighting on multiple-choice test. *Journal of Educational Measurement*, 8, 291-296.

- Henriksson, W. (1981). *Uppgiftsformat och partiell kunskap på ett kvantitativt numeriskt prov* (Pedagogiska rapporter, nr 97). Umeå universitet.
- Henriksson, W.; Lexelius, A. & Hamrén, M. (1985). *Försök med ett nytt numerisk prov* (Prov memoria, nr 6). Umeå Universitet.
- Koehler, R.A. (1971). A Comparison of the validities of conventional choice testing and various confidence marking procedures. *Journal of Educational Measurement*, 8, 297-303.
- Rippey, R.M. (1968). Probabilistic Testing. *Journal of Educational Measurement*, 5, 211-215.
- Rippey, R.M. (1970). A comparison of five different scoring functions for confidence tests. *Journal of Educational Measurement*, 7, 165-170.
- Stanley, J.C. & Wang, M.D. (1968). *Differential weighting. A survey of methods and empirical studies*. New York: College Entrance Examination Board, 1968.
- Wang, M.D. & Stanley, J.C. (1970). Weighting test items and test-item options. An overview of the analytical and empirical literature. *Educational and Psychological Measurement*, 30, 21-35.
- Wedman, I. (1969). *Poängsättning av samtliga svarsalternativ i multiple-choice uppgifter och dess effekter på reliabiliteten* (Pedagogiska rapporter, nr 2). Umeå universitet.
- Wedman, I. (1970). *Olika poängsättningsteknikers effekter på reliabiliteten vid bedömning av flervalfrågor* (Pedagogiska rapporter, nr 12) Umeå universitet.