

PISA 2012 OG JAKTEN PÅ DEN EFFEKTIVE REKTOREN

*EN ANALYSE AV LEDELSESKONSTRUKTER
OG DERES SAMMENHENG MED ELEVENES
RESULTATER I NORGE OG SVERIGE*

OLE ANDREAS NORDHAUG



MASTEROPPGAVE I
UTDANNINGSLEDELSE (UV4090) VED
INSTITUTT FOR LÆRERUTDANNING OG SKOLEFORSKNING

UNIVERSITETET I OSLO

NOVEMBER 2014

PISA 2012 JAKTEN & PÅ DEN EFFEKTIVE REKTOREN

EN ANALYSE AV LEDELSESKONSTRUKTUR
OG DERES SAMMENHENG MED ELEVENES
RESULTATER I NORGE OG SVERIGE

OLE ANDREAS NORDHAUG

© Ole Andreas Nordhaug

2014

PISA 2012 og jakten på den effektive rektoren

Ole Andreas Nordhaug

<http://www.duo.uio.no/>

Trykk: Reprosentralen, Universitetet i Oslo

IV

Sammendrag

PISA-undersøkelsene har det siste tiåret vært toneangivende for nordisk utdanningspolitikk og deler av utdanningsforskningen, og offentliggjøringen av elevenes resultater møtes med stor interesse av media. Det som derimot er blitt viet langt mindre oppmerksomhet, er dataene som samles inn fra skolenes rektorer. Disse bærer i seg et betydelig potensiale med tanke på kvantitativ og komparativ rektorforskning generelt, og forskning på effektiv ledelse spesielt.

Denne studien tar utgangspunkt i ledelseskonstruktene som inngår i PISA 2012 sin Index of Principal Management, og undersøker om disse har en begrepsvaliditet som er god nok til at de kan brukes for å belyse relevante trekk ved norske og svenske rektors pedagogiske ledelse av skolen. De statistiske og teoretiske analysene som benyttes i denne sammenheng, inngår også i en prosess der svarene fra de norske og svenske rektorene tilknyttet rektors pedagogiske ledelse undersøkes på nytt. Formålet med dette er å etablere alternative ledelseskonstrukt som er bedre egnet til å belyse likheter og ulikheter mellom norske og svenske rektorer, og som ikke minst lar seg bruke i strukturelle modeller der sammenhengen mellom rektors pedagogiske ledelse og elevenes resultater undersøkes.

Presentasjonen av norsk og svensk skole i teoridelen fokuserer på hvordan endringene i landenes utdanningssystemer de siste tiårene har påvirket grunnskolerektorenes mandat og handlingsrom. Når de ulike ledelseskonstruktene analyseres, gjøres dette med utgangspunkt i Hallinger (2003) sin fremstilling av instructional og transformational leadership, som er de to lederstilene som i størst grad har preget forskningsfeltet tilknyttet effektiv ledelse. De to lederstilene trekkes også inn som en sentral del av analyseapparatet i forbindelse med sammenligningen av svarene fra de norske og svenske rektorene.

Studiens funn og deres betydning for forskningsfeltet kan kort oppsummeres slik:

- De statistiske og teoretiske analysene av PISA sin Index of Principal Management viste at de fire konstruktene som inngår i indeksen har betydelige svakheter, både når det kommer til konstruktens psykometriske egenskaper og dere teoretiske fundament. Hvis det er et mål at PISA-undersøkelsene i fremtiden skal kunne gi flere og bedre analysemuligheter tilknyttet rektors pedagogiske ledelse, vil det være hensiktsmessig å både videreutvikle og utvide denne delen av skole spørreskjemaet.

- Jakten på alternative ledelseskonstrukturer viste seg å føre frem, og de nye konstruktene danner ikke bare et utgangspunkt for en sammenligning av noen sentrale trekk ved norske og svenske rektorers pedagogiske ledelse, men gjør det også mulig å undersøke om det finnes signifikante sammenhenger mellom de ulike ledelseskonstruktene og elevenes resultater i PISA 2012. De alternative konstruktene kan også fungere som et konkret innspill i forbindelse med videreutviklingen av PISAs ledelseskonstrukturer.
- Sammenligningen av svarene fra de norske og svenske rektorene tilknyttet de alternative ledelseskonstruktene viste at rektorene i de to landene, i likhet med utdanningssystemene forøvrig, utviser flere likhetstrekk enn forskjeller. Likevel ser man at svenske rektorer i noe større grad enn sine norske kolleger vektlegger ledelsesfunksjoner som forbindes med instructional leadership, der rektor inntar en mer aktiv rolle, og i større grad styrer, kontrollerer og evaluerer skolens praksis.
- Funnene tilknyttet de strukturelle modellene viser at det i mange tilfeller ikke lar seg gjøre å påvise en signifikant sammenheng mellom rektors ledelse og elevenes resultater. I de tilfellene der sammenhengen er signifikant, er regresjonskoeffisientene svært svake, noe som bidrar til å kaste et kritisk lys over betydningen av rektors pedagogiske ledelse for elevenes resultater i norske og svenske skoler i dag.

Forord

1 mastergrad.

2 glimrende veiledere.

3 neglisjerte familiemedlemmer.

4 år med utdanningsledelse.

3: Nora, Iver og Alma: Er snart tilbake for fullt.

2: Christian Brandmo og Rolf Vegar Olsen: Takk.

1: God fornøyelse.

Innholdsfortegnelse

1	INNLEDNING	1
1.1	Introduksjon.....	1
1.2	Problemstilling og forskningsspørsmål	2
1.3	Forskningsdesign	2
1.4	Studiens muligheter og begrensninger	3
1.5	Oversikt over kapitlene.....	4
2	TEORI.....	5
2.1	Rektors mandat og handlingsrom	5
2.2	Forskning på effektiv ledelse.....	12
2.3	Etablering av begrepsvaliditet	18
3	METODE	22
3.1	PISA 2012.....	22
3.2	Analyser utført med Mplus7	27
3.3	Analyser utført med SPSS 21	30
3.4	Forskningsetikk	32
4	RESULTATER	33
4.1	Analyse av PISA-konstruktene.....	33
4.2	Etablering av alternative konstrukter.....	35
4.3	Likheter og forskjeller mellom Norge og Sverige.....	44
4.4	Rektors ledelse og elevenes læring.....	47
5	DISKUSJON	51
5.1	Begrepsvaliditeten til PISA-konstruktene	51
5.2	Alternative ledelseskonstrukter i PISA 2012.....	57
5.3	Likheter og forskjeller mellom Norge og Sverige.....	61
5.4	Pedagogisk ledelse og elevenes resultater	66
5.5	Validitet	69
5.6	Studiens bidrag til forskningsfeltet.....	72
6	LITTERATUR	75
7	VEDLEGG	82
7.1	Item tilknyttet pedagogisk ledelse i PISA 2012	82
7.2	Item tilknyttet “School management” i TALIS 2008.....	83

1 INNLEDNING

1.1 Introduksjon

Nordisk forskning på skoleledelse har tradisjonelt, som Møller (2006a) påpeker, ønsket å ”forstå ledelse som fenomen mer enn å identifisere kjennetegn ved god ledelse” (s. 37). Forskningsfeltet har vært preget av kvalitative studier som har gitt mange verdifulle bidrag, men som i liten grad har vært generaliserbare på nasjonalt nivå, ettersom den ytre validiteten har vært hemmet av for små utvalg. I internasjonal sammenheng, og spesielt i amerikansk skoleforskning, har man derimot siden 70-tallet sett mange eksempler på kvantitative studier av såkalte ”effektive skoler”. Her har man nettopp forsøkt å identifisere det vi kan kalle ”kjennetegn ved god ledelse”, og spesielt de siste tre tiårene har man fått en utvidet forståelse for det komplekse samspillet mellom rektors pedagogiske ledelse og elevenes læring (Hallinger, 2003, 2011; Hallinger og Heck, 2010; Robinson, 2010; Robinson, Lloyd, og Rowe, 2008). Møller (2006b) er klar på at ”den amerikanske forskningen om skoleledelse har vært totalt dominerende på feltet”, og konkluderer med at ”det trengs utvilsomt studier som tematiserer skoleledelse i norsk kontekst” (s. 105). Dette ståstedet finner vi også igjen hos Johansson og Bredeson (2011), som i en analyse av fremtidige forskningsperspektiv på rektor, blant annet fremhever behovet for kvantitative rektorstudier, herunder behovet for studier som kan undersøke sammenhengen mellom rektor og elevens resultater (s. 72).

Offentliggjøringen av resultatene fra PISA-undersøkelsene, som gjennomføres i regi av OECD, er en stor begivenhet i nordisk skoledebatt, og elevenes resultater danner grunnlag for et skred av avisartikler og debattinnlegg. Resultatene fra rektorenes spørreskjema, som blant annet undersøker sider ved rektors pedagogiske ledelse av skolen, er derimot lite brukt utover enkle deskriptive analyser. Gjennom å analysere begrepsvaliditeten til ledelseskonstruktene som benyttes i PISA 2012, har denne studien som mål å gi et bidrag til å styrke dette feltet i nordisk sammenheng med tanke på fortolkning av eksisterende så vel som fremtidige data. En analyse av eksisterende og alternative ledelseskonstrukt som tar utgangspunkt i rektorenes svar, kan videre benyttes i sekundære analyser av mer substansiell art, blant annet gjennom å undersøke sammenhengen mellom rektors pedagogiske ledelse og elevenes resultater.

I denne studien brukes begrepet *konstrukter* gjennomgående, da dette benyttes i det meste av litteraturen tilknyttet PISA, også her til lands. Det refereres i denne studien likevel til *begrepsvaliditet*, da dette er langt mer utbredt enn det tilsvarende *konstruktvaliditet*.

1.2 Problemstilling og forskningsspørsmål

Studien har som mål å besvare følgende problemstilling:

Er ledelseskonstruktene tilknyttet Index of School Management i PISA 2012 egnet til å beskrive relevante trekk ved norske og svenske rektorers pedagogiske ledelse, herunder sammenhengen mellom rektors pedagogiske ledelse og elevenes resultater, og finnes det eventuelt et grunnlag for alternative konstrukter som kan belyse dette bedre?"

Dette vil gjøres gjennom å fokusere på følgende forskningsspørsmål:

- Har konstruktene som inngår i PISA 2012 sin Index of Principal Management en tilfredsstillende begrepsvaliditet sett i lys av svarene til norske og svenske rektorer?
- Kan man, basert på itemene tilknyttet pedagogisk ledelse i PISA 2012, etablere alternative ledelseskonstrukter med en mer tilfredsstillende begrepsvaliditet?
- Hvilke likheter og forskjeller mellom rektors mandat og handlingsrom i Norge og Sverige kan ledelseskonstruktene bidra til å belyse?
- Kan man, ved å bruke strukturell modellering, finne en signifikant sammenheng mellom ledelseskonstruktene og elevenes resultater?

1.3 Forskningsdesign

Denne studien tar utgangspunkt i PISA 2012, som er en kvantitativ tverrsnittsundersøkelse. I motsetning til det meste av PISA-forskning forøvrig, vil studien fokusere mer på rektorene og deres svar tilknyttet pedagogisk ledelse enn på elevene og deres resultater, selv om disse vil bli benyttet i studiens siste fase. Studien er i det store og hele tredelt, der første fase har som mål å vurdere begrepsvaliditeten til ledelseskonstruktene som inngår i PISA 2012 sin Index of Principal Management, heretter referert til som *PISA-konstruktene*. Dette gjøres med bakgrunn i psykometriske analyser, samt ved å drøfte hvordan konstruktene forholder seg til relevant teori om effektiv ledelse. I neste fase foretas sekundærundersøkelser av dataene fra de to landene. Formålet med dette er å etablere alternative ledelseskonstrukter som samsvarer bedre med svarene fra de norske og svenske rektorene. Her vil det benyttes ulike multivariate

analysemetoder, men både den psykometriske og den innholdsmessige delen av begrepsvaliditeten til de nye konstruktene må til syvende og sist vurderes. Det er et mål at de alternative konstruktene skal være identiske for begge land, slik PISA-konstruktene er felles for alle OECD-landene, da dette utgjør grunnlaget for å kunne sammenligne de to landene. Konstruktene som viser høyest grad av begrepsvaliditet skal i siste fase danne utgangspunkt for strukturelle modeller som undersøker sammenhengen mellom rektors pedagogiske ledelse og elevenes resultater innen de tre områdene regning, lesing og naturfag.

1.4 Studiens muligheter og begrensninger

Mulighetene for å analysere data fra rektorer og elever sammen er et viktig aspekt ved PISA-undersøkelsene som i liten grad har vært utnyttet til nå. Blant annet gjelder dette muligheten for strukturelle modeller som undersøker sammenhenger mellom rektors ledelse og elevenes resultater. En enkeltstående tverrsnittsundersøkelse som dette vil naturlig nok aldri kunne påvise kausale sammenhenger, da dette forutsetter en eksperimentell tilnærming som er svært vanskelig å få gjennomført innen utdanningsforskning. Det som derimot er mulig her, er å generalisere funnene opp på nasjonalt nivå, ettersom elevene som deltok fra Norge og Sverige utgjør et representativt utvalg i sine respektive land. PISA sin funksjon som en komparativ undersøkelse kjenner vi godt igjen fra rangeringslistene norske elever figurerer på med jevne mellomrom. Den byr riktignok også på muligheten for å sammenligne trekk ved rektorenes hverdag og deres pedagogiske ledelse av skolen som, i tillegg til å ha verdi i seg selv, kan bidra til en dypere forståelse av studiens funn. Det å bruke allerede eksisterende data legger nødvendigvis også noen begrensninger på studien, spesielt med tanke på at man til enhver tid er prisgitt de begrepsoperasjonaliseringene av rektors pedagogiske ledelse som allerede finnes i spørreskjemaene. Dette innebærer ikke nødvendigvis at man er bundet av de ledelseskonstruktene som ligger til grunn for itemene, men man mister muligheten til å kvalitetssikre itemene og etter eget ønske utforme item som undersøker de aspektene ved rektors pedagogiske ledelse man er mest interessert i. Som ved alle storstilte kvantitative undersøkelser, ofres de rike beskrivelsene og muligheten for å stille oppfølgingsspørsmål som intervju eller andre kvalitative metoder kan gi. PISA-undersøkelsene mangler i tillegg lærernivået på den enkelte skolen, som kunne vært nyttig for å kontrollere rektorenes rapportering av sin egen pedagogiske ledelse.

1.5 Oversikt over kapitlene

KAPITTEL 2 redegjør for de begreper, teorier og kontekster som er nødvendige for å angripe problemstillingen. Her vil rektors mandat og handlingsrom i Norge og Sverige presenteres og sammenlignes, i tillegg til at utvalgte sider ved forskning på effektiv ledelse gjennomgås. Til slutt vil det etableres et teoretisk rammeverk for begrepsvalideringen i studien som dekker både de eksisterende og de alternative ledelseskonstruktene.

KAPITTEL 3 har som mål å sette leseren inn i de metodiske vurderingene som ble gjort i forbindelse med den psykometriske delen av begrepsvalideringen, samt redegjøre for prosessen som leder frem mot etableringen av de alternative konstruktene. Kapitlet avsluttes med en redegjørelse for de forskningsetiske implikasjonene ved studien.

KAPITTEL 4 følger firedelingen fra forskningsspørsmålene, der første del vil vise resultatene fra den psykometriske delen av begrepsvalideringen tilknyttet PISA-konstruktene. Neste del har som mål å presentere veien frem mot de alternative ledelseskonstruktene, deres psykometriske egenskaper og relevante likheter og ulikheter tilknyttet de norske og svenske rektorenes svar. Deretter vil svarene fra de norske og svenske rektorene sammenlignes, før den siste delen av kapitlet presenteres de strukturelle modellene som undersøker om det finnes signifikante sammenhenger mellom rektors pedagogiske ledelse og elevenes resultater.

KAPITTEL 5 inneholder en utdypende diskusjon av funnene i studien, sett i lys av relevant teori om norske og svenske rektors mandat og handlingsrom, teori tilknyttet effektiv ledelse og de grunnleggende prinsippene for begrepsvalidering og den metodiske tilnærmingen som helhet. Avslutningsvis vil styrker og svakheter ved studiens indre og ytre validitet drøftes, før dens bidrag til forskningsfeltet og mulige fremtidige studier vil bli presentert.

2 TEORI

Teorikapittelet av studien har som mål å presentere teori og begrep innen skoleledelse, forskning på skoleledelse og begrepsvaliditet. Den første delen skal kaste lys over hva som kjennetegner rektorrollen i Norge og Sverige, og skal først og fremst utgjøre konteksten som funnene i resultatdelen skal drøftes innenfor, selv om den også skal vise at det finnes en balanse mellom likhetstrekk og relevante forskjeller mellom landene som gjør det interessant å sammenligne dem. Den andre delen av kapittelet vil fokusere på forskning på effektiv ledelse. Det vil her kort redegjøres for hvordan forskningsfeltet har utviklet seg, og de viktigste kjennetegnene ved to sentrale teorier vil bli presentert. Disse to skal utgjøre ryggmargen i den innholdsmessige delen av begrepsvalideringen, før forskningsfeltets posisjon i Norden oppsummeres til slutt. Den siste delen av kapittelet vil ta for seg teori tilknyttet validitet, mer spesielt hvordan begrepsvaliditet kan undersøkes og etableres, som vil danne utgangspunktet for den metodiske tilnærmingen.

2.1 Rektors mandat og handlingsrom

En forståelse av rektors mandat og handlingsrom i Norge og Sverige er nødvendig for å kunne diskutere likheter og ulikheter i funnene tilknyttet rektorenes pedagogiske ledelse i de to landene. Gjennomgangen vil ta for seg noen sentrale områder for utvikling i norsk og svensk skole fra 70-tallet og frem til i dag, og har som mål å belyse hvilke konsekvenser utviklingen har hatt for rektorrollen. Rundt 1970 var situasjonen i begge land preget av at rektorene så godt som alltid var rekruttert fra lærerkollegiet, og rektors rolle var primært preget av administrasjon og byråkrati (Imsen, 2004, s. 144; Ullman, 1997, s. 177). Sejersted (2005) slår videre fast at utdanningspolitikken frem til dette handlet mer om politikk enn om pedagogikk, og at dette var likt i både Norge og Sverige. Fokuset lå på en sentralstyrt enhetsskole, der mer skole, og ikke nødvendigvis en bedre skole, var nøkkelen til kunnskapsrike og produktive medborgere (s. 296ff). På 80- og 90-tallet skal vi se at *desentralisering* og *New Public Management* (NPM) utgjør de viktigste trendene, før OECD, *nyliberalisme* og en *ny statlig styring* gjør seg gjeldende i begge land på 2000-tallet. På tross av mange fellestrekk divergerer likevel norsk og svensk skole noe gjennom hele denne perioden, og det vil avslutningsvis pekes på noen utvalgte likheter og forskjeller mellom grunnskoleledere i Norge og Sverige som er spesielt relevante med tanke på å tolke resultatene i denne studien.

2.1.1 Desentralisering og deregulering

På slutten av 60-tallet og utover 70-tallet sprer det seg en motvilje mot den sentral- og detaljstyrte skolen, og rektor blir trukket frem i lyset som en motvekt til dette. En vektlegging av lokalt spillerom og handlekraft sprer seg fra USA til Europa, og senere også til Norden, og idealrektorens rolle beveger seg fra administrator og implementør til inspirator og innovatør. (Helgøy, 2006; Imsen, 2004; Møller, Ottesen, Dyrdal Solbrekke, og Aaserud, 2000). I Norge får dette få konkrete konsekvenser på 70-tallet, men Møller et al. (2000) trekker frem tre viktige trekk ved norsk skole på 80-tallet som bidro til en desentralisering av utdanningssystemet her til lands. For det første flyttes, med nedleggningen av Forsøksrådet, ansvaret for utviklingsarbeidet i skolen over på lokalt nivå. Dette følges, for det andre, opp i 1986 av en ny ordning med blokkfinansiering av kommunene, som fjerner de øremerkede tilskuddene og gir kommunene større makt i sine økonomiske prioriteringer, inkludert i spørsmål som angår skolen. For det tredje erstattes M74 med en ny mønsterplan (M87), som gir skolene mer frihet til å velge ut lærestoff tilpasset de lokale forholdene (s. 210f), selv om synet på reelle effekten av dette er delt (Aasen et al., 2012, s. 42).

Også i Sverige signaliserer staten utover 70-tallet at den ønsker å gå over til et mer desentralisert og målstyrt system (Nygren og Johansson, 2000, s. 263). Det innføres i denne sammenheng et friere finansieringssystem i 1978, og nye læreplaner i 1980, som i likhet med M87 ga mer frihet lokalt (Lundahl, 2002, s. 689ff). Dereguleringen har som mål å effektivisere skolen (Forsberg, 2010, s. 467; SOU, 2014:5, s. 57), men selv om desentraliseringsprosessene på 80-tallet var omfattende, var det like fullt bare starten på en prosess som for alvor setter fart på 90-tallet (Forsberg, 2010; Lundahl, 2002; Nygren og Johansson, 2000; Richardson, 2004; SOU, 2014:5). I 1990 blir kommunene hovedansvarlige for driften av skolene, herunder arbeidsgiveransvaret, og fra 1991 får kommunene i enda større grad styre over midlene de får fra staten (Forsberg, 2010, s. 467; Helgøy, 2006, s. 100). Den økonomiske omleggingen bærer riktignok også i seg at skolene må bli mer kostnadseffektive, og skolens innhold blir delvis lidende når fokuset må flyttes over på organisasjons- og budsjettspørsmål (Nihlfors, 2010, s. 500f).

2.1.2 Målstyring og ansvarliggjøring

Med 90-tallet gjorde NPM og målstyringsprinsipper for alvor sitt inntog i norsk offentlig sfære, mye på grunn av OECD-evalueringen av norsk skole i 1988 (Aasen et al., 2012;

Langfeldt, 2008; Møller og Skedsmo, 2013; Skedsmo, 2011). Helgøy og Homme (2006) fremhever i denne sammenheng spesielt *Lov om kommuner og fylkeskommuner* fra 1992, fordi lokale myndigheter nå fikk ansvaret for ikke bare å tilby utdanning, men også å sikre kvaliteten på denne (s. 146). Møller et al. (2000) mener likevel at tiltakene som skulle sikre lokale myndigheter og skolene stor frihet, ble kraftig hemmet av detaljerte læreplaner og stram økonomi i kommunene (s. 212), og L97 bryter tydelig med desentraliseringen vi ellers ser i skolen ved å fjerne mange av de lokale tilpasningsmulighetene som fantes i M87 (Helgøy og Homme, 2006, s. 146f; Imsen og Volckmar, 2014, s. 40). Selv om OECD allerede i 1988 hadde påpekt et behov for en større grad av kontroll av skolens resultater (Skedsmo, 2011, s. 76), skal man først på 2000-tallet se fruktene av denne prosessen. Norges svake PISA-resultater blir en katalysator for dette (Lillejord, 2011, s. 291; Møller og Skedsmo, 2013, s. 343), og når Kvalitetsutvalgets rapport, NOU 2002:10 (2002), nedsetter rammene for et nasjonalt vurderingssystem, får dette raskt den tilslutningen som behøves, og ordningen vedtas av Stortinget i 2004. Dette skiftet beskrives av Skedsmo (2011) som ”et skifte i styringen av norsk skole, fra hovedsakelig bruk av mål- og innholdsverktøy til økt grad av resultatstyring” (s.76). Blant verktøyene i det nye nasjonale vurderingssystemet finner vi de nasjonale prøvene. Beslutningen om at resultatene skal offentliggjøres, fører til at rektor ikke bare gjøres ansvarlig for skolens resultater innad i utdanningssystemet, men også i pressen. Motstanden mot prøvene, og spesielt offentliggjøringen av disse, brer seg i skolen, og Engeland og Langfeldt (2009) rapporterer om at mange rektorer opplever prøvene som et uttrykk for ”en manglende tillit til rektorenes ansvarlige ledelse av egen skole” (s. 12).

NPM gjør seg gjeldende også i Sverige utover 90-tallet (Lundahl, Erixon Arreman, Holm, og Lundström, 2013, s. 499; SOU, 2014:5, s. 110), noe som blir spesielt synlig med Lpo94, som for alvor innfører mål- og resultatstyring i skolen. Begrunnelse for dette er at lærere og rektorer lokalt har de beste forutsetningene for å vite hvordan de nasjonale målene best kan nås (SOU, 2014:5, s. 65). I denne forbindelse, og hele åtte år tidligere enn i Norge, innføres det i 1996 nasjonale prøver, og rektors ansvar for skolens resultater betones i større og større grad (SOU, 2014:5, s. 198). Prøvene har siden den gang økt i omfang, og Sverige er per i dag et av OECD-landene med mest bruk av slike prøver (OECD, 2013d, s. 158). Rankinglister av skoler basert på elevenes prestasjoner i ulike prøver og tester møter også i Sverige kritikk fra flere hold, og Forsberg (2010) trekker frem dette som et lysende eksempel på en reduksjon av skolens oppdrag, kunnskapen den skal formidle og dannelsesperspektivet i skolen (s. 472).

2.1.3 Friskoler og markedsstyring

Friskoleloven av 2003 gjorde det enklere å opprette private skoler her til lands, og loven representerte samtidig en innføring av markedsstyring av skolen, som i følge Langfeldt (2008) ”utgjør et grunnleggende brudd med den statlige styringen som til nå har regulert sektoren” (s. 96). Gjeninnføringen i 2007 av kravet om at friskolene skulle utgjøre et religiøst eller pedagogisk alternativ, legger imidlertid en demper på nyetablering av friskoler, og alt i alt går fortsatt 97,5 % av norske elever i dag ved offentlige skoler¹, og på tross av at muligheten er til stede, er det fortsatt få kommuner som har innført fritt skolevalg der pengene følger eleven.

I Sverige ser vi derimot allerede på 80-tallet at samtidens individualistiske tidsånd fører til et krav om retten til å starte opp og velge skole, og så tidlig som i 1994 innføres det fritt skolevalg med stykkprisfinansiering (Lundahl, 2002, s. 692). Mens 1,0% av elevene går i private grunnskoler i 1992, er tallet 13,7% i 2014², og man kan i dag i realiteten snakke om et parallelt skolesystem i områdene der privatiseringen har kommet lengst. Lundahl et al. (2013) konkluderer med at markedstankegangen i dag har fått et solid fotfeste i Sverige. (s. 511).

2.1.4 Rektorutdanning og resentralisering

På tross av flere tiltak som skal styrke rektorene faglig, er det først med St.meld. nr. 30 (2003-2004) og innføringen av Kunnskapsløftet at vi i Norge for alvor ser konturene av en ”lederutdanning på masternivå for skoleledere og lærere”, der målet er at ”alle rektorer skal få et tilbud om slik lederopplæring” (s. 100f). Ordningen skal styrke skolene som ”lærende organisasjoner” i lys av studier ”som belyser hva som kjennetegner skoleledelsen ved utviklingsorienterte skoler som kan dokumentere godt læringsutbytte” (s. 29). Den påfølgende etableringen av en formalisert skolelederutdanning og et nasjonalt kvalitetssystem er eksempler på tiltak som ”ettertrykkelig plasserte ansvaret for skolens resultater hos rektor, som dermed må bli mer opptatt av kvaliteten på de pedagogiske prosessene” (Lillejord, 2011, s. 293). Hybertsen Lysø et al. (2012) viser riktignok at få av rektorene som deltar har en forventning om å endre lederstil, en forventning som også innfris, og at de i stedet vektlegger å få økt trygghet i rollen (s. 66). Alle de nasjonale tiltakene som nevnes her, samt innføringen av statlige tilsyn i 2008, kan ses på som et steg i retning av en gradvis resentralisering i skolen, der staten gjennom ulike styringsverktøy søker å øke sin innflytelse.

¹ http://www.utdanningsforbundet.no/upload/Publikasjoner/Faktaark/Faktaark%202014/Faktaark_2014.01.pdf

² <http://www.ekonomifakta.se/sv/Fakta/Valfarden-i-privat-regi/Skolan-i-privat-regi/Elever-i-friskola/>

Det er tydelig at svensk skole har hatt et sterkere fokus på rektors utdanning som et ledd i utviklingsprosessene i skolen, og har også en sterkere sentral organisering enn den norske (Hybertsen Lysø, Stenaker, Aamodt, og Mjøen, 2011, s. 38ff). Selv om fokuset på 80- og 90-tallet fortsatt ligger på administrative oppgaver, viser evalueringene likevel at rektorer som har deltatt blir noe mer pedagogisk innrettet i sitt arbeid (Nygren og Johansson, 2000, s. 265f). I tråd med endringene i skolen på 90-tallet, endres også rektorutdannelsen gradvis fra og med 2001, med økt vektlegging av skolejuss og myndighetsutøvelse, mål- og resultatstyring og skolelederskap (Forsberg, 2010, s. 460; Nygren og Johansson, 2000, s. 290f). Omfanget av det nasjonale tilsynet med skolen er også kommet lenger i Sverige enn i Norge, og SOU (2014:5) fremhever tilsynsordningen som et tegn på en resentralisering, der statens myndighet nå ikke lenger stopper ved kommunegrensene. Ved bruk av store ressurser og et omfattende mandat, driver Skolinspektionen med tilsyn av alle Sveriges kommuner og skoler. Målet er å sørge for at den kommunale styringen av skolene planlegges, gjennomføres og dokumenteres på tilfredsstillende vis, og tidligere kritikk av et for dyrt og omfattende byråkrati synes å ha havnet i glemmeboka. (Forsberg, 2010, s. 471).

2.1.5 Rektors mandat og handlingsrom

Selv om kravet om formell lærerutdanning for landets rektorer i 2004 fjernes fra opplæringsloven i Norge, er nær sagt alle rektorer også i dag kvalifiserte som lærere gjennom en lærerutdanning (jfr. Aasen et al., 2012, s. 51). Møller (2011) mener dette kommer til syne ved at undersøkelsene og kartleggingene ved skolene har påfallende få konsekvenser for skolens virksomhet (s. 45). Møller (2011) finner også, med bakgrunn i blant annet TALIS-undersøkelsen fra 2008, at graden av reell ansvarliggjøring, både fra skoleeier overfor rektor og rektor overfor lærerne, later til å tilpasse seg den allerede etablerte praksisen i skolen. Møller (2011) bruker her begrepet ”skjult kontrakt” (s. 44), som vi kjenner fra Berg (1990), og det er interessant at vi finner igjen flere de samme konklusjonene både hos Møller et al. (2000) og Møller (2011). Dette kan tyde på at endringene i rektors mandat og handlingsrom i praksis har vært mindre omfattende enn omveltningene i skolen ellers skulle tilsi.

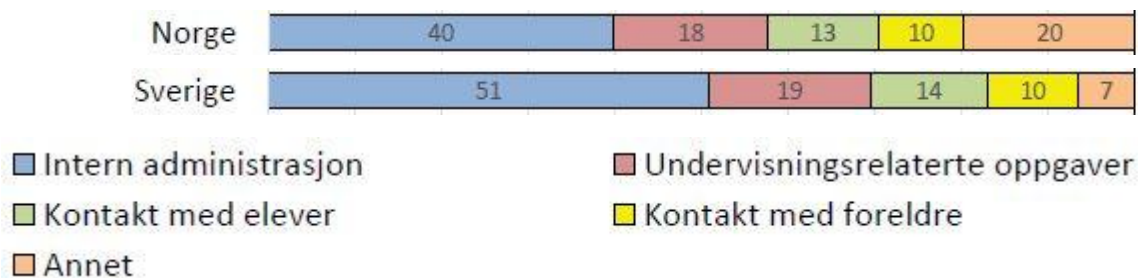
Pierre (2007) konkluderer med at desentraliseringen i Sverige på 90-tallet begrenset den politiske styringen av skolene til fordel for en mer selvstendig styring fra rektor (Nihlfors, 2010, s. 505). Det råder riktignok noe uenighet rundt omfanget av dette, og i Skolverket (1998) kan vi lese at det, selv om rektorene er positive til reformene i utdanningssystemet,

ikke lar seg påvise noen reell forandring i skolens ledelse. Det påpekes også at rektor fortsatt bruker mye tid på administrative oppgaver (s. 55f). Dette er trekk vi kjenner igjen fra beskrivelsen av Norge før tusenårsskiftet, og ”den skjulte kontrakten” gjør seg, i følge Skolverket (1998), fortsatt gjeldende, nesten ti år etter Berg (1990), gjennom en mangel på ledelse av og støtte til lærerne fra rektors side (s. 33f). Et trekk som derimot har preget Sverige i større grad enn Norge de siste tiårene, er privatiseringsprosessene som blant annet har bidratt til at svenske rektorer i dag har et større handlingsrom enn sine norske kolleger, spesielt i de private skolene (Lundahl et al., 2013; OECD, 2002, 2008a, 2014a). Privatiseringen har også ført til at rektorer i dag føler et press fra andre skoler i nærmiljøet i kampen om elevene, og har fått en rolle som nærmer seg en bedriftsleder, der strategisk tankegang og brukernes ønsker spiller en større rolle enn tidligere (Lundahl et al., 2013, s. 506). På tross av disse forskjellene, peker Johansson og Bredeson (2011) på at rektorenes manglende vilje til å sette seg inn i det som skjer i klasserommene gjør seg gjeldende også i Sverige, og at de sterke normene tilknyttet lærernes profesjonelle autonomi, hindrer rektorene i å få et innblikk i klasserommet og påvirke pedagogiske beslutninger (s. 65f).

2.1.6 Hvordan egner landene seg for sammenligning?

Helgøy og Homme (2006) konkluderer med at Norge og Sverige, på tross av å tilhøre samme velferdsmodell, utviser noen nevneverdige forskjeller i det som ellers er to sammenfallende utdanningssystemer, som understreker at konvergens på noen områder ikke betyr at det er slik over det hele (Helgøy og Homme, 2006, s. 142). Når Thuen (2010) eksempelvis konkluderer med at det er for tidlig å avgjøre om enhetsskolen her til lands har blitt erstattet av markedstenkning og NPM (s. 85), er det for Sveriges del er en utbredt enighet om at man i dag langt på vei har fått en markeds- og resultatstyrt skole (Lundahl et al., 2013; SOU, 2014:5). Dette har blant annet ført til at svenske rektorer er noe mer autonome enn sine norske kolleger (Hansen et al., 2014; Haug, 2010; Helgøy og Homme, 2004, 2006, 2007), noe som også bekreftes i en rekke OECD-undersøkelser (OECD, 2008a, 2008b, 2010a, 2014b).

Den siste OECD-undersøkelsen, TALIS 2013, er noe mer omfattende i deres kartlegging av grunnskolerektorer og deres hverdag enn PISA-undersøkelsene. Resultatene fra de tilhørende rapportene inneholder noen sammenligninger som det kan være interessant å inkludere her, da de kan berike diskusjonen i denne studien ytterligere. Et eksempel på dette er Figur 2-1 på neste side, som viser hva grunnskolerektorer oppgir at de bruker tiden sin på.



Figur 2-1: Prosentvis fordeling på seks aktiviteter for skoleledere på ungdomstrinnet – TALIS 2013
 Utdrag av Figur 3.2 (Carlsten, Caspersen, Vibe, og Aamodt, 2014, s. 25)

Som vi ser, er det store likhetstrekk mellom de kategoriene som kan ha en tilknytning til rektors pedagogisk ledelse, som ”Undervisningsrelaterte oppgaver”, ”Kontakt med elever” og ”Kontakt med foreldre”. Men likhetene i TALIS 2013 stopper ikke der, og omhandler også forhold som ikke angår rektors pedagogiske ledelse direkte. Carlsten et al. (2014) viser eksempelvis til at norske og svenske grunnskoleledere fortsatt rekrutteres primært fra lærerstanden, at de har en tilsvarende kjønnsfordeling og aldersfordeling, og at de samlet sett har jobbet omtrent like lenge som rektor (s. 66 og 72). Det er viktig å ha disse likhetene i bakhodet i denne studien, da eventuelle forskjeller som påvises i denne studien følgelig ikke kan avskrives ved å peke på markante forskjeller i sammensetningen av rektorstanden. Dette tyder på at forskjellene vi finner mellom Norge og Sverige i TALIS 2013 og PISA 2012 i stedet kan tenkes å henge sammen med ulikhetene i landenes utdanningssystemer. I tillegg til rektors autonomi, som allerede er nevnt, ser vi også en noe høyere grad av ledelsesutdanning i Sverige. Her har omtrent 80% av rektorene ”sterk grad” av ledelsesutdanning, mens tallet for Norge ligger på om lag 65% (s. 70). Andelen rektorer som ikke har deltatt i utviklingsaktiviteter det siste året er også større i Norge med 10%, mot 4% i Sverige (s. 20).

Hvis landene skal kunne sammenlignes med utgangspunkt i ulike ledelseskonstrukt, er det også nødvendig at disse konstruktene er de samme for begge landene. Konstruktene som inngår i PISA sin Index of Principal Management, benyttes naturlig nok som utgangspunkt i alle deltakerlandene, men i denne studien skal det også etableres alternative konstrukt som skal brukes i begge land. Disse konstruktene skal baseres på eksplorative faktoranalyser gjennomført på hvert enkelt land, som så sammenlignes i håp om at det er mulig å kombinere dem i et felles sett med alternative ledelseskonstrukt. I denne sammenheng er det en styrke at de semantiske forskjellene i oversettelsen, med tanke på setningsbygning og begrepsbruk, vil være små mellom de to landene, da dette minsker faren for at itemene tolkes ulikt. For ordens skyld bør det her påpekes at kun den norske utgaven av rektorspørreskjemaet er analysert i denne studien.

2.2 Forskning på effektiv ledelse

Sammenhengen mellom rektors ledelse og studentenes læring var lenge ansett som selvfølgelig, men på 60-tallet blir det publisert flere studier som spenner ben under denne antakelsen. Disse studiene, deriblant Benjamin Blooms *Stability and Change in Human Characteristics* fra 1964 og James Colemans *Equality of Education Opportunity* fra 1966, viser at elevenes sosiale bakgrunn og andre faktorer som ligger utenfor skolen, er langt viktigere for elevens prestasjoner enn de skolebaserte faktorene, herunder rektors ledelse. Mot slutten av 60-tallet og i de påfølgende tiår blir det satt i gang en rekke studier, spesielt i USA, som undersøker sammenhengen mellom skole og elevresultater nærmere, og som indirekte søker å gjenopprette troen på skolen som relevant for elevens læring (Sergiovanni, 2009, s. 195). Med dette er jakten på den ”effektive” rektoren i gang, en rektor som gjennom sin ledelse skal ha en målbar effekt på elevenes læring. Denne perioden gir opphav til en lang rekke av det man kan kalle kjennetegn på effektiv ledelse, som samles i ulike lederstiler, herunder *instructional leadership*, *transformational leadership*, *distributed leadership* og *shared leadership*. Mot slutten av 70-tallet viser spesielt studier av fattige storbykoler en klar sammenheng mellom styrkede elevresultater og en tydelig pedagogisk ledelse, ofte kalt *instructional leadership*. Denne lederstilen får følgelig stor innflytelse utover 80-tallet i USA, men resultatene viser seg i stor grad å være et biprodukt av studiene, og tvetydigheten tilknyttet rektors rolle som bidragsyter i forbindelse med skoleutvikling består (Hallinger, 2011, s. 127). Utover 80- og 90-tallet tar derfor en rekke forskere på seg oppgaven med å undersøke effekten av de ulike lederstilene, ikke bare *instructional leadership*. Selv om flere studier støtter opp under en positiv sammenheng mellom ulike kjennetegn på effektiv ledelse, er det stemmer som er kritiske til validiteten til disse studiene. Studier utenfor USA som ønsker å finne de samme positive effektene, blant annet i Nederland, finner få signifikante effekter av rektors ledelse, og ved en nærmere undersøkelse av studier i USA og Storbritannia viser det seg at resultatene heller ikke her er like samstemte som retorikken skulle tilsi (Hallinger og Heck, 1998, s. 158).

Utover 2000-tallet styrkes riktignok rektors rolle som en signifikant aktør med tanke på å utvikle, påvirke og opprettholde skoleutviklingsinitiativer med positive læringsresultater innen forskningsfeltet. Selv om OECD (2008c) konkluderer med at det fortsatt ikke finnes evidens for noen sammenheng mellom lederstil og elevenes resultater (s. 200), kommer Robinson et al. (2008) frem til et annet resultat. Her sammenfattes 27 studier som ser på

sammenhengen mellom ulike lederstiler og elevresultater, og metastudien viser en gjennomsnittlig effektstørrelse (z-score) for instructional leadership på 0.42, som innebærer at gruppen ligger 0.42 standardenheter over kontrollgruppen. Den tilsvarende effektstørrelsen for transformational leadership ligger på 0.11 og ”andre” ligger på 0.30 (s. 657). På tross av funnene i denne studien, knytter det seg fortsatt en betydelig usikkerhet til hva som er de egentlige årsakene, og hvor godt de lar seg overføre mellom ulike kontekster, ikke minst på tvers av landegrenser (Jacobson og Ylimaki, 2011, s. 1). Hallinger (2011) peker også på at debatten det siste tiåret har vært mindre opptatt av hvilken av de ulike lederstilene som best forklarer sammenhengen mellom ledelse og elevresultater, ettersom de empiriske resultatene har stabilisert seg. I stedet ser man nå at elementer fra de ulike lederstilene kombineres på nye måter og danner grunnlag for nye modeller (s. 127).

I jakten på et relevant teoretisk rammeverk er det naturlig å undersøke hva rammeverket tilknyttet PISA 2012 selv har å komme med på dette punktet, men det viste seg at det her står svært lite om det teoretiske grunnlaget for konstruktene. Noen av itemene til de fire ledelseskonstruktene i PISA 2012 finnes delvis igjen i PISA 2009 (se item i *kursiv* i vedlegg 7.1), men heller ikke i rammeverket tilknyttet PISA 2009 er det enkelt å oppdrive noe konkret tilknyttet dette. Det var mer eller mindre tilfeldig at en forbindelse til itemene i PISA 2012 ble oppdaget i rammeverket til TALIS 2008 (OECD, 2010b, s. 34f), og denne forbindelsen ble via epost bekreftet fra sentralt hold i PISA-ledelsen i OECD. Rammeverket viser blant annet til at flere av konstruktene i TALIS 2008 baserer seg på Hallingers *Principal Instructional Management Rating Scale* (PIMRS) (OECD, 2010b, s. 34). Rammeverket henviser riktignok også til OECD (2008b), *Improving School Leadership, Volume 1*, som en kilde til det teoretiske grunnlaget for konstruktene, men mangelen på noen nærmere henvisning til hva bidraget herfra har vært, gjør det noe vanskelig å følge tråden videre, da det i OECD (2008b) refereres videre til en hel rekke bidragsytere. Bruken av PIMRS er et viktig argument for at denne studien har valgt å ta utgangspunkt i instructional leadership og Hallinger (2003), der essensen ved denne lederstilen, og herunder PIMRS, blir oversiktlig gjengitt. Artikkelen påpeker videre at transformational leadership tradisjonelt har utgjort et analytisk motstykke til instructional leadership (s. 329), og lederstilene blir sammenlignet på en oversiktiglig måte, gjengitt under. Denne studien vil, på bakgrunn av dette, basere de teoretiske analysene på Hallinger (2003) og det bildet artikkelen gir av instructional og transformational leadership.

2.2.1 Instructional Leadership (IL)

Instructional leadership er som nevnt et begrep som har vært mye brukt i internasjonal forskning på effektive skoler, men har fått lite innpass i forskning på nordiske rektorer. Hvis man er ute etter definisjoner på IL må man da også gå til nordamerikansk litteratur. Southworth (2002) trekker frem tre reviewartikler som spesielt viktige. Den første, Leithwood, Jantzi, og Steinbach (1999) definerer at IL "Assumes that the critical focus for attention by leaders is the behaviors of teachers as they engage in activities directly affecting the growth of students" (s. 8). Den andre artikkelen, Blase og Blase (1998) mener på sin side at den viktigste praksisen innen IL består i å jobbe for å styrke lærernes profesjonelle utvikling (s. 11), noe som skal gjøres gjennom samtaler, å fostre refleksjon, være synlige, belønne gode resultater og fremheve lærernes autonomi (s. 18). Den siste artikkelen Southwood fremhever er Hallinger og Heck (1998), der de bygger videre på en modell utviklet av Hallinger og Murphy (1985). Denne modellen finner vi også igjen i Hallinger (2003), og har vært toneangivende i det siste tiårs debatt innen feltet, og som her skal danne et teoretisk bakteppe for analysene av ledelseskonstruktene i PISA 2012.

Hallinger (2003) sammenfatter IL i tre dimensjoner, som til sammen innbefatter ti funksjoner som rektor skal fylle gjennom sin praksis: Den første dimensjonen fokuserer på *skolens mål*, og går ut på at rektor skal lede arbeidet med å (1) defineres skolens visjon og formulere målbare mål og (2) sørge for å gjøre disse kjent blant kollegiet. Den andre dimensjonen heter *ledelse av undervisning*, og handler om kontroll og læreplanarbeid. Her skal rektor (3) kontrollere og evaluere undervisningen, (4) koordinere læreplanarbeidet og (5) evaluere elevenes resultater og fremgang. Den tredje dimensjonen dreier seg om hvordan rektor sørger for å *styrke skolen som lærende organisasjon*, blant annet gjennom høye standarder og forventninger og kultur for læring. Dette innebærer å (6) sette av tid til arbeidet, (7) fremme faglig utvikling, (8) være synlig gjennom hele prosessen, (9) skape insentiver for lærerne og (10) insentiver for læring (s. 355). Alle disse finner vi igjen i PIMRS, som også har med en ekstra ledelsesfunksjon, "Utvikle akademiske standarder" (Hallinger, 1994, s. 2ff).

2.2.2 Transformational Leadership (TL)

Transformational leadership utvikler seg på 70- og 80-tallet som en teori for lederskap på generell basis, og får spesielt på 90-tallet tilhengere som hadde sett seg lei på lederskap dominert av en ovenfra-og-ned-tankegang, som blant annet kjennetegnet IL på den tiden.

Bernhard Bass, som bidro sterkt til å videreutvikle denne lederstilen, fremhever at TL er mer egnet enn topptunge teorier i kontekster som er avhengige av tillit mellom leder og underordnet og der selvtillit, motivasjon og mening hos den enkelte er viktig (Bass, 1997, s. 173). At dette slår an i skolekretser er neppe overraskende, da lærerkollegiet er et av kroneksempelene på det Bolman og Deal (2005) henviser til som et ”fagbyråkrati”, preget av stor grad av autonomi og en sterk profesjonsforståelse. Fagbyråkratier som organisasjonsform har mange fordeler, men kan by på utfordringer med tanke på samordning og kvalitetskontroll, samt at det som system er lite mottakelig for endringer (s. 101). Bolman og Deal (2005) peker også på at endringen man har erfart i utdanningskretser, mot et system som flytter ansvaret for undervisningen høyere opp i systemet, har møtt mye motbør. Utviklingen fortsetter utover 90-tallet med bidrag fra blant annet Leithwood og Jantzi (2000), som henter Bass’ tankegods over i utdannelses-sammenheng. Deres modell brukes av Hallinger (2003) som utgangspunkt for hans analyser, og vil følgelig fylle samme funksjon i denne studien.

Modellen består av åtte komponenter som skal inngå i endringsarbeidet, fordelt på fire nivåer: *Det grunnleggende nivået* fokuserer på viktigheten av (1) individuell støtte, der rektor skal sørge for at alle lærerne får en god og personlig tilpasset oppfølging. *Det andre nivået* går på at rektor skal sørge for at organisasjonen har (2) en tydelig visjon som utvikles til (3) samsvarende mål i kollegiet. *Det tredje nivået* tydeliggjør at rektor ikke driver endringen direkte, men i stedet skal (4) sørge for intellektuell stimulans og (5) skape kultur for læring. *Det siste nivået* understreker at modellen er avhengig av (6) rektor som forbilde, (7) høye forventninger og (8) belønningssystemer for å fungere.

2.2.3 Likheter og forskjeller mellom IL og TL

Hallinger (2003) anerkjenner at de to lederstilene har mange likhetstrekk og områder der de i stor grad overlapper hverandre, men viser også til områder der de to modellene skiller seg fra hverandre. Likhetene og ulikhetene tydeliggjøres ved at rektors ulike ledelsesfunksjoner innen instructional og transformational leadership sammenstilles og sammenlignes i tabellen som gjengis på neste side:

INSTRUCTIONAL LEADERSHIP (IL)	TRANSFORMATIONAL LEADERSHIP (TL)	LIKHETER OG ULIKHETER MELLOM IL OG TL
Utvikle og spre skolens visjon og målsetninger innad i personalet	Tydelig visjon og felles mål for skolen	IL ser på skolens mål som overordnet, mens TL vektlegger sammenhengen mellom skolens målsetninger og personalets personlige mål.
Kontroll med læreplanarbeid, tilsyn og evaluering av undervisning, evaluering av elevprestasjoner og prioritering av tid til undervisning.	---	Det finnes ingen tilsvarende elementer for disse kontroll- og koordinasjonsfunksjonene i TL, da TL bærer i seg en antakelse om at disse funksjonene vil ivaretas av andre i organisasjonen.
---	Individuell støtte	IL legger til grunn at dette vil ivaretas gjennom tilsyn og læreplankoordinering. TL ser på arbeid mot ansattes individuelle mål som helt sentralt.
Høye forventninger og insentiver for lærere	Høye forventninger og belønningssystemer	Tilsvarende fokus på belønning som følger opp skolens ønskede utvikling.
Sørge for profesjonell utvikling av personalet	Intellektuell stimulans	IL har fokus på profesjonell utvikling i samsvar med skolens mål, mens TL også er opptatt av de personlige målene, som ikke trenger være nært tilknyttet skolens overordnede mål.
Høy grad av synlighet	Modellering	Tilsvarende funksjoner, der rektor skal ha en tydelig rolle med samsvar mellom liv og lære.
---	Utvikle skolekultur	Dette er også et fokus i IL, men skal ivaretas av andre dimensjoner i lederstilen.

Tabell 2-1: Likheter og forskjeller mellom IL og TL - fritt etter Hallinger (2003) (s. 344)

Grovt sammenfattet kan man si at begge lederstilene er opptatt av klare og tydelig kommuniserte mål, selv om TL vektlegger de ansattes personlige mål i tillegg til skolens overordnede mål. Begge lederstilene er opptatt av høye forventninger til de ansatte og belønningssystemer, og de fremhever begge rektors rolle som forbilde. Når det gjelder ulikhetene, vil en IL-rektor være mer direkte involvert i læreplanarbeid, ha et større fokus på bruk av ulike kontrollmekanismer og vil være mer opptatt av elevresultatene i evalueringsarbeidet. En TL-rektor vil på sin side være mer opptatt av at det utvikles en skolekultur der den ansattes personlige mål er vel så viktige som skolens mål, og der rektor skal tilrettelegge for utvikling og bidra med individuell støtte. Det er også viktig å poengtere at TL generelt sett fokuserer langt mindre på rektors eksplisitte rolle i endringsarbeidet. Selv om rektor selvfølgelig alltid vil være sentral i en slik prosess, kan store deler av ansvaret delegeres bort til andre i lederteamet, så vel som til lærere (s. 342ff).

2.2.4 Forskning på effektiv ledelse i Norden

Instructional og transformational leadership er lite utbredt som begreper i norsk skoleforskning, og mangelen på etablerte norske oversettelser av begrepene bærer vitne om at nordisk forskning har vært opptatt av å ”forstå ledelse som fenomen mer enn å identifisere kjennetegn ved god ledelse” (Møller, 2006a, s. 37). Som vi har sett, har dette lenge vært et gjennomgående trekk ved nordisk forskning på effektiv ledelse, og studier av effektiv ledelse slik vi kjenner det fra nordamerikansk forskning er følgelig nærmest ikke-eksisterende. Møller (2006a) vier riktignok noe oppmerksomhet til de to nevnte lederstilene, og sammenfatter disse kjapt med at begge har likhetstrekk med begrepet ”pedagogisk ledelse”, men at IL anses for å være ”mer preget av didaktikk og klasseromspraksis og samtidig mer ovenfra-og-ned-orientert”, mens TL er ”en type endringsledelse som transformerer organisasjonen til et høyere nivå både med tanke på motivasjon og moralsk handling” (s. 32).

Møller (2006b) peker på at interessen for effektiv ledelse har økt de siste årene (s. 99), men at det er foretatt få studier i Norden. Studiene som finnes, viser at sammenhengen i beste fall er indirekte, og konkluderer ofte med at gode skoler kjennetegnes av et:

”tydelig og demografisk, men kraftfullt lederskap fra rektors side som fremfor alt er innrettet mot skolens kunnskapsmål, samarbeid mellom lærerne om undervisningen, høye forventninger til elever og medarbeiderne, og etablering av prosedyrer for belønning for godt utført arbeid.” (s. 104)

Sammenligner vi dette med Tabell 2-1 ser vi at dette sammenfaller nokså godt med kjennetegn på både IL og TL, selv om fokuset på en kraftfull leder og skolens kunnskapsmål muligens heller noe mer i retning av IL. Emstad og Postholm (2010) representerer også et eksempel på den fornyede interessen for effektiv ledelse, da de har forsøkt å bringe IL over i norsk kontekst. De oversetter IL med ”FUL-ledelse”, der FUL står for *Forbedring av undervisning og læring*, og deres redegjørelse for hva IL og TL representerer, sammenfaller med Hallinger (2003). Caspersen (2011) er et annet, og kanskje spesielt interessant tilfelle, fordi studien faktisk prøver å se på sammenhengen mellom skolekultur, herunder rektors ledelse, og elevenes resultater. Caspersen (2011) baserer seg på TALIS 2008, som vi jo vet deler sitt teoretiske rammeverk med PISA 2012. Analysene av rektors ledelse settes riktignok ikke direkte i sammenheng med elevenes resultater, og de signifikante funnene i studien dreier seg heller om variasjon innad på en skole enn mellom skolene (s. 85ff).

I denne studien er det derimot et mål å se etter signifikante sammenhenger mellom ulike ledelseskonstrukter og elevenes resultater, men da er det avgjørende at disse konstruktene har en akseptabel, og helst en god begrepsvaliditet. Hvordan en slik begrepsvalidering kan og bør undersøkes, er også neste post på programmet.

2.3 Etablering av begrepsvaliditet

Et avgjørende element i denne studien dreier seg om begrepsvalidering av ulike ledelseskonstrukter. I første omgang dreier det seg om konstruktene tilknyttet PISA 2012 sin Index of School Management, men senere også alternative konstrukter som kan samsvare bedre med rektorenes faktiske svar. Det er viktig å påpeke at mye av litteraturen tilhørende begrepsvaliditet dreier seg om prosessen med å lage og gjennomføre undersøkelser på en forsvarlig måte, slik at både indre og ytre validitet ivaretas. Da denne studien tar utgangspunkt i allerede gjennomførte undersøkelser, er de aspektene ved validiteten som ligger forut for selve undersøkelsen utenfor studiens kontroll. Ulempene dette medfører er allerede nevnt, og denne delen vil i stedet fokusere på ulike sider ved det å undersøke og etablere begrepsvaliditet i etterkant av undersøkelsen, og indre og ytre validitet vil kun forklares kort.

2.3.1 Indre og ytre validitet

Cohen, Manion, Morrison, og Bell (2011) omtaler både indre og ytre validitet. For det første, definerer de at *indre validitet* “seeks to demonstrate that the explanation of a particular [...] issue [...] which a piece of research provides can actually be sustained by the data” (s. 183). Kort sagt dreier dette seg om at dataene som ligger til grunn for studien støtter opp under og bidrar til å belyse temaet som behandles og de tilhørende funnene. *Ytre validitet* defineres på sin side som ”the degree to which the results can be generalized to the wider population, cases, settings, times and situations” (s. 186). Her er altså sakens kjerne hvorvidt funnene har relevans ut over de personene som har deltatt i studien, det spesifikke tidspunktet for undersøkelsen osv., kort sagt funnenes overførbarhet. Begrepsvaliditet forholder seg til både indre og ytre validitet, fordi god eller dårlig begrepsvaliditet påvirker både hvor relevante funnene er med tanke på å beskrive temaet, og følgelig funnenes overførbarhet.

2.3.2 Begrepsvaliditet

Cook og Beckman (2006) er ofte brukt som et utgangspunkt for argumentasjon tilknyttet begrepsvaliditet. Her samles ulike innfallsvinkler under samlebetegnelsen *construct validity*, og det listes opp ulike kilder som kan bidra til å styrke begrepsvaliditeten til et konstrukt. Av de fem innfallsvinklene artikkelen nevner; *innhold*, *svarprosess*, *indre struktur*, *relasjon til andre variabler* og *konsekvenser* (s. 9ff), er to ikke tilgjengelige for denne studien. Dette gjelder *relasjon til andre variabler*, som omhandler korrelasjon til andre testverktøy/metoder som tester samme konstrukt, samt en innfallsvinkel som kalles *konsekvenser*. Denne siste er mest aktuell i forbindelse med medisinske studier, der det styrker validiteten om konsekvensene av tiltak basert på studien er i tråd med studiens funn. Vi står derfor igjen med tre kilder til begrepsvaliditet, *innhold* (content), *svarprosess* (response process) og *indre struktur* (internal structure), som nå vil presenteres.

INNHold

Også når det gjelder innholdsdimensjonen tilknyttet begrepsvalideringen møter man i denne studien noen begrensninger. Det har nemlig vist seg å være vanskelig å finne god informasjon om det Cook og Beckman (2006) fremholder som den viktigste bestanddelen tilknyttet innholdsvalideringen av konstruktene, nemlig *construct definition*, som beskrives som "the instruments intended purpose, the process for developing and selecting items, the wording of individual items, and the qualifications of the item writers and reviewers" (s. 10). En del av de tekniske sidene ved dette er redegjort for i forbindelse med PISA 2009 (OECD, 2012), men bortsett fra at ordeggingen til itemene i skolespørreskjemaet er tilgjengelig, sier rammeverket til PISA 2012 i praksis ingenting om prosessen bak begrepsoperasjonaliseringene. Vi kan riktignok finne noe i rammeverket til TALIS 2008, som PISA 2012 henter noen av sine item fra, men som vi skal se i diskusjonsdelen, er også dette nokså begrenset. Innholdsdelen av begrepsvaliditeten koker derfor ned til en analyse av begrepsoperasjonaliseringen til konstruktene og de tilhørende itemene slik de står, noe som i denne studien skal gjøres med utgangspunkt i sentral teori tilknyttet effektiv ledelse, og spesielt instructional og transformational leadership, som ble presentert i forrige del av teorikapittelet.

SVARPROSESS

Det meste som inngår i denne delen av begrepsvalideringen dreier seg om ulike prosedyrer og verktøy som kan brukes for å få kunnskap om tankeprosessene til den enkelte respondenten mens spørreskjemaet besvares, samt hvordan data håndteres og sikres (Cook og Beckman,

2006, s. 10). Disse er ikke offentliggjort for PISA 2012 enda, men rutinene som ble benyttet ved PISA 2009 kan finnes i den tekniske rapporten, OECD (2012), og viser med all tydelighet hvor omfattende dette arbeidet er. Detaljene tilknyttet dette vil ikke beskrives ytterligere, da det blir for omfattende, samt at dette er et område der PISA er anerkjent for å ha gode rutiner (OECD, 2011). En annen side ved svarprosessen, hvordan itemene scores, vil derimot bli berørt i metodekapittelet. For denne studiens del er dette primært aktuelt i forbindelse med Likert-skalaen som rektorene bruker for å angi hyppigheten av de ulike ledelsesaktivitetene, mens det for elevenes del kan være et tema i forbindelse med bruken av plausible values, som også vil gjennomgå i metodekapittelet.

INDRE STRUKTUR

Cook og Beckman (2006) nevner at både reliabilitetsanalyser og faktoranalyser er verdifulle verktøy når man ser etter indre struktur og konsistens innenfor konstruktene, men vier lite oppmerksomhet til faktoranalyser ut over dette. Når det gjelder reliabilitet, finnes flere måter å vurdere dette på, men en del av disse, som test-retestprosedyrer er ikke mulige innenfor denne studien. I stedet er det naturlig å bruke Cronbachs α , og en nærmere beskrivelse av hva dette innebærer kommer i metoddelen. Det er likevel viktig å ta med seg videre at reliabilitet alene aldri vil være tilstrekkelig for å påvise begrepsvaliditet (Cook og Beckman, 2006, s. 13).

Når det gjelder bruken av konfirmerende og eksplorerende faktoranalyser i forbindelse med begrepsvalidering, vil denne studien også benytte seg av Miller, McIntire, og Lovler (2011), da de behandler dette noe grundigere, og også viser til studier der konfirmerende faktoranalyser eller en kombinasjon av eksplorerende og konfirmerende faktoranalyser er brukt i forbindelse med begrepsvalidering (s. 255f). Her pekes det på at faktoranalyser er velegnet for å vise at item som tilhører samme konstrukt, og som derfor forventes å korrelere med hverandre, faktisk gjør dette, og at de på den andre side ikke korrelerer med item som tilhører andre konstrukt. Både eksplorerende og konfirmerende faktoranalyser kan fortelle oss dette gjennom å angi hvor sterkt itemene lader mot konstruktene de er ment å tilhøre, og at de samtidig lader svakt mot de andre konstruktene. Det er derimot kun konfirmerende faktoranalyser som kan gi oss et mål for modelltilpasning, det vil si hvor godt den teoretisk baserte modellen passer overens med de empiriske observasjonene (s. 258). Konfirmerende faktoranalyser brukes primært når det allerede eksisterer et sett med konstrukt som skal testes på et dypere nivå enn det en ren reliabilitetsanalyse kan gjøre, og denne metoden vil følgelig benyttes på konstruktene i PISA sin Index of School Management.

Når det gjelder bruken av eksplorerende faktoranalyser, viser Miller et al. (2011) til en studie der man på tross av å ha en forventning om hvilke konstrukter som var til stede, først gjennomførte en eksplorerende faktoranalyse. Konstruktene som kom til syne i denne analysen, ble deretter bekreftet med en konfirmerende faktoranalyse. I denne studien kan de eksplorerende faktoranalysene i seg selv bidra til å belyse likheter og forskjeller mellom landene, men primært er likevel målet å lage et felles sett med alternative konstrukter på bakgrunn av disse faktoranalysene. Disse må så gjennom en psykometrisk og innholdsmessig valideringsprosess gjennom en reliabilitetsanalyse, en konfirmerende faktoranalyse og en diskusjon av begrepsoperasjonaliseringen. De metodiske valgene som er foretatt i forbindelse med de statistiske analysene vil nå bli presentert.

3 METODE

Ettersom hele denne studien hviler på holdbarheten til de kvantitative analysene som foretas, følger det at redegjørelsen for de metodiske valgene som er gjort i den forbindelse må være tilsvarende omfattende. Studiens forskningsdesign er som nevnt tredelt, men tar som helhet utgangspunkt i PISA 2012, og første del av metodekapittelet vil gjøre rede for relevante sider ved forskningsdesignet i PISA 2012, herunder hvilke hensyn som må tas når man skal analysere dataene fra undersøkelsen. Deretter vil programvaren som er benyttet for å utføre de ulike statistiske analysene bli presentert, og det vil gjøres rede for de metodiske valgene som er tatt i forbindelse med disse. De statistiske begrepene som er sentrale for å forstå resultatene, og den påfølgende diskusjonen, blir forklart fortløpende gjennom kapittelet.

3.1 PISA 2012

OECDs Programme for International Student Assessment (PISA) er primært kjent fordi den måler 15-åringers kompetanse innenfor fagområdene lesing, matematikk og naturfag på et tidspunkt som i de fleste land representerer avslutningen av den obligatoriske skolegangen (Kjærnsli og Olsen, 2013a, s. 13). PISA-undersøkelsene er alle kvantitative undersøkelser, der informasjonen innhentes ved hjelp av ulike spørreskjema og oppgaver med kvantifiserbare resultater (OECD, 2013a). Ettersom PISA gjennomføres hvert tredje år, kan landene som deltar spore endringer over tid, selv om det naturlig nok ikke er de samme elevene som deltar hver gang, noe som regnes som en trendstudie (Cohen et al., 2011, s. 266). Denne studiens fokus vil derimot kun ligge på PISA 2012, der 196 norske og 207 svenske skoler deltok, og vil følgelig behandles som en enkeltstående studie. PISA 2012 er nemlig den første og hittil eneste PISA-undersøkelse der det er gjort et forsøk på å dele inn pedagogisk ledelse i navngitte ledelseskonstrukter. Selv om PISA 2009 hadde flere av de samme itemene tilknyttet rektors pedagogiske ledelse som vi finner igjen i PISA 2012, og selv om man også i PISA 2009 ser konturene av ledelseskonstruktene som beskrives i PISA 2012, behandles de her kun under ett, i det som kalles *Index of principal's leadership* (OECD, 2010a, s. 99).

Ettersom datainnsamlingen for PISA 2012 skjer innenfor en kort, begrenset tidsramme, defineres den som en tverrsnittsstudie (Cohen et al., 2011, s. 267). PISA faller også inn under det Kleven (2002) definerer som et ikke-eksperimentelt, eller deskriptivt design, ettersom det ikke er noen uavhengige variabler som manipuleres i designet. Dette er typisk for

utdanningsforskningen generelt, ettersom bevisst manipulasjon både har praktiske og etiske utfordringer tilknyttet seg, og statistiske sammenhenger vil derfor som regel aldri være forenelig med flere kausalrelasjoner. Jakten på alternative tolkninger og skjulte variabler vil i slike tilfeller være sentralt med tanke på å styrke den indre validiteten, og multippel regresjon og strukturell modellering kan bidra til å styrke den statistiske kontrollen. Disse analysene er riktignok begrenset til de observerte variablene som allerede er en del av analysen, noe som innebærer at også variablenes teoretiske fundament vurderes nøye (s. 265ff).

Både analysene av PISA-konstruktene og de alternative konstruktene i denne studien bærer i seg dette todelte perspektivet, noe som altså fordrer både en innholdsmessig gjennomgang av itemene og psykometriske analyser av de innsamlede dataene. Det teoretiske bakteppet som skal danne grunnlaget for denne analysen er allerede presentert, men de sentrale metodiske aspektene tilknyttet studiens kvantitative analyser må fortsatt gjennomgås nærmere.

3.1.1 Hvordan undersøkes rektorenes pedagogiske ledelse?

Det er i alt 21 item i skole spørreskjemaet (se vedlegg 7.1) der rektor skal angi hvor ofte han utfører visse ledelsesoppgaver tilknyttet den pedagogiske ledelsen av skolen. Hvis man ser bort fra de semantiske forskjeller som oppstår når disse itemene oversettes til ulike språk, er de 21 itemene like for alle deltakerlandene. Roe og Kjærnsli (2013) er tydelige på at en internasjonal komparativ undersøkelse som dette vil gi en større variasjon innen de kartlagte aspektene enn en særnorsk undersøkelse vil kunne gi, men sier videre at dette i prinsippet gir oss "et datasett med stort analytisk potensiale – både for å beskrive hva skoleledelse er i en norsk kontekst, og hvor et eventuelt potensiale for å utvikle lederrollen ligger" (s. 238). 13 av de 21 itemene er tilknyttet fire forhåndsdefinerte ledelseskonstrukter som går under paraplybenevnelsen *Index of School Management*. Disse er *Framing and communicating the school's goals and curricular development* (LEADCOM), *Instructional leadership* (LEADINST), *Promoting instructional improvements and professional development* (LEADPD), og *Teacher participation* (LEADTCH) (OECD, 2013b, s. 140). Det er viktig å påpeke at disse konstruktene er latente variabler, eller faktorer. De er med andre ord ikke direkte observert, og utledes hver for seg på bakgrunn av tre til fire item i skole spørreskjemaet som er ment å ta for seg ulike aspekter ved konstruktet. Rektorene besvarer itemene ved å angi med hvilken frekvens de er involvert de ulike ledelsesoppgavene innenfor seks forhåndsdefinerte svarkategorier: "Aldri", "1-2 ganger i året", "3-4 ganger i året", "Én gang i

måneden”, ”Én gang i uken” og ”Mer enn én gang i uken”. Kategoriene er innbyrdes ekskluderende og kan rangeres etter størrelse, men avstanden mellom hver svarkategori er ikke identisk, noe som betyr at vi befinner vi oss på ordinale målenivå. Mer spesifikt er skalaen som benyttes her en variant av det som ofte refereres til som Likert-skalaer (Cohen et al., 2011, s. 386). Det å analysere slike skalaer parametrisert er et omstridt fenomen, ettersom det strengt tatt kreves en intervallskala med identisk avstand mellom de ulike svarkategoriene, noe som i verste fall kan gå ut over svarprosess-dimensjonen ved begrepsvalideringen. Kuzon, Urbanchek, og McCabe (1996) redegjør for hvorfor man på prinsipielt grunnlag ikke bør ta seg slike friheter, men denne praksisen har likevel blitt mer og mer utbredt, spesielt i samfunnsvitenskapelige surveystudier, de siste tiårene. Carifio og Perla (2007) viser at man, ved å behandle itemene gruppevis og ved å ha et tilstrekkelig antall svarkategorier, minst fem og helst syv, i de fleste tilfellene vil kunne trekke pålitelige konklusjoner også når skalaene behandles parametrisert (s. 115). Denne konklusjonen støttes av Norman (2010), som understreker at dette også gjelder ved små utvalg, ujevn varians og ikke-normalitet (s. 631). På bakgrunn av dette går det i denne studien god for å behandle rektordataene parametrisert.

3.1.2 Hvordan trekkes skolene og rektorene?

For å representere variasjonen i populasjonen best mulig, vil et utvalg av 4000 tilfeldig valgte elever alltid være bedre enn for eksempel 40 elever fra 100 skoler (OECD, 2009, s. 39), men det å gjennomføre prøven med 4000 tilfeldige elever er av økonomiske og praktiske årsaker ikke mulig. Det trekkes derfor ut grupper av elever ved et utvalg skoler, og både i Norge og i Sverige har det vært et mål at samplingen av skolene, og følgelig elevene, skal danne et utvalg som, så godt det lar seg gjøre, representerer den nasjonale populasjonen av elever. I metodevedlegget til Kjærnsli og Olsen (2013b) kan vi lese at sannsynligheten for at en skole skulle bli trukket, og følgelig denne skolens rektor, var proporsjonal med hvor mange av landets 15-åringer som gikk på denne skolen. Den svenske hovedrapporten har ikke et like detaljert metodevedlegg, og selv om man her kan man lese at ”Syftet var at ge et representativt urval av skolar som har elever födda 1996” (Fredriksson et al., 2013, s. 22), må dette forstås i lys av at rektorene ikke er representative i seg selv, men i kraft av elevene de representerer (OECD, 2009, s. 146). Det å gjøre analyser med rektordataene fra de 196 norske og 207 svenske rektorene direkte, tilsier derfor også at resultatene må tolkes ut ifra at rektorene ikke er representative i seg selv, men at vi her har et utvalg av rektorer for et representativt utvalg elever. Strengt tatt bør man derfor i alle henvisninger til rektorene i PISA

omtale rektorene i kraft av elevene de representerer, som for eksempel ”20% av norske elever har en rektor som aldri gjennomfører uformelle observasjoner i klasserommene”. I det følgende vil derimot rektorene, som hos Roe og Kjærnsli (2013), likevel omtales direkte, slik at eksempelet over i stedet vil lyde ”20% av norske rektorer gjennomfører aldri uformelle observasjoner i klasserommene”.

3.1.3 Hvordan trekkes og vektes elevene?

Når en elev ved en skole trekkes, blir det automatisk trukket ut 30 15-åringer fra denne skolen. Hvis antall 15-åringer er lavere enn 30, deltar samtlige elever. På tross av at sannsynligheten for at skolen trekkes er proporsjonal med antall elever, kan det være forhold som gjør at hvert elevcase må vektes ut ifra hvor mange elever denne representerer i den nasjonale populasjonen. Slik anerkjenner man at noen caser er viktigere enn andre med tanke på hvilket bidrag de tilfører den nasjonale populasjonen, noe som reduserer muligheten for bias i dataene (OECD, 2009, s. 48). OECD (2009) nevner i denne sammenheng at vektingen vil være spesielt viktig hvis noen deler av populasjonen av nasjonale hensyn er bevisst over- eller underrepresentert (s. 36), men i både den norske og den svenske hovedrapporten legges det vekt på at utvalget skal være representativt, og Kjærnsli og Olsen (2013b) konkluderer også med at resultatene for Norges del blir nokså like med vektete og uvektete data (s. 316). På bakgrunn av dette er det rimelig å anta at de vektete og uvektete resultatene vil bli nokså like også for Sveriges del, men dette er ikke eksplisitt nevnt i den svenske rapporten. Selv om vektingen på grunnleggende nivå neppe vil være utslagsgivende ved enkle beregninger, er det et annet aspekt ved dette som fortsatt bør beskrives, nemlig replicate weights.

3.1.4 Hva menes med Replicate Weights?

Som nevnt trekkes elevene av praktiske hensyn ut i grupper, og man kan av naturlige årsaker forvente at elever ved samme skole har en del samsvarende erfaringer. Elevenes svar innad på en skole samsvarer derfor som regel mer med hverandre enn med svarene fra andre skoler. Dette kan føre til at noen item får en kunstig lav spredning, som igjen kan gi for lave standardfeil og feilmarginer (Kjærnsli og Olsen, 2013b, s. 315). Faren med dette er at sammenhenger som egentlig ikke er signifikante kan oppgis som signifikante, noe som vil svekke validiteten til hele undersøkelsen. For å kompensere for dette, benytter PISA seg av en metode som kalles *Balanced Repeated Replication method* (BRR), og mer spesifikt en variant

av dette kalt *Fay's modification* (OECD, 2009, s. 70). Enkelt forklart medfører dette at det genereres 80 ekstra datasett med *replicate weights* (RW), der halvparten av skolene i hvert av datasettene trekkes tilfeldig ut og får sine vekter redusert til 0,5 mens den andre halvparten får vektene økt til 1,5 (OECD, 2009, s. 73). Beregningen av standardfeil gjøres først med den originale vektene og deretter med de 80 replikatene, og variasjonen i disse beregningene gir oss et nytt og bedre mål for variansen, som igjen vil minske sannsynligheten for at ikke-signifikante sammenhenger oppgis som signifikante. OECD (2009) anbefaler at man alltid bruker RW når man skal rapportere standardavvik og signifikans (s. 46), selv om det påpekes at dette problemet er minst i land med små forskjeller mellom skolene (s. 63). For Norges del ligger svært lite av den totale variansen på skolenivå, men ettersom man i den norske rapporten har tatt høyde for dette fenomenet ved å bruke RW (Kjærnsli og Olsen, 2013b, s. 315), vil dette også benyttes i denne studien.

3.1.5 Hva er Plausible Values?

Hvis vi kun anerkjenner elevens scorer innen de tre fagområdene som punkttestimater når man gjør beregninger, gir dette et for snevert syn på hva denne eleven bidrar med til helheten, da hver case rent metodisk representerer en gruppe tilsvarende elever. I stedet for å bruke elevens score direkte, trekkes det derfor i stedet fem scorer, såkalte *Plausible Values* (PV), fra en sannsynlighetskurve rundt eleven, hvor det naturligvis vil være en større sannsynlighet for at disse ligger nær toppen av elevens kurve (OECD, 2009, s. 94ff). Wu (2005) beskriver dette godt: "One way to describe plausible values is to say that plausible values represent the range of abilities that a student might possibly have, given the students item responses" (s. 115). I PISA angis eksempelvis elevens prestasjon innen matematikk som PV1MATH t.o.m. PV5MATH, og alle disse skal ideelt sett brukes når det skal gjøres beregninger som skal generaliseres til populasjonen, da studier har vist at denne metoden gir mindre bias enn ved bruk av punkttestimatene (Von Davier, Gonzalez, og Mislevy, 2009; Wu, 2005). OECD (2009) viser riktignok at forskjellen på å bruke kun én av PVene kontra fem er svært små ved større utvalg, og konkluderer med at det ikke har noen praktisk betydning om man bruker kun én eller alle PVene ved utvalg på størrelse med dem vi finner i denne studien (s. 45f). Fordi det å kombinere RW og PV i én og samme strukturell modell har vist seg å være svært komplisert i Mplus, tar denne studien utgangspunkt i én PV, da dette bør ha en minimal effekt på resultatene. Mer spesifikt brukes kun PV1MATH, PV1READ og PV1SCIE, som for enkelthets skyld angis som MATH, READ og SCIE i modellene i resultatdelen.

3.2 Analyser utført med Mplus7

I denne studien brukes programmet Mplus7 av Muthen og Muthen til analysene som faller inn under det som kalles *structural equation modelling* (SEM). Konfirmerende faktoranalyser (CFA) er en variant av dette, mens sammenhengen mellom ledelseskonstruktene og elevenes resultater er en mer renyrket SEM. I denne delen vil disse analysemetodene presenteres, samtidig som det redegjøres for relevante metodiske valg som er av betydning for modellene.

3.2.1 Hva er SEM, og hvordan er dette blitt håndtert i studien?

SEM betegner en familie med statistiske prosedyrer som har blitt mer og mer utbredt, spesielt innen samfunnsvitenskapen. SEM brukes når man ønsker å innhente informasjon om konstrukter som ikke lar seg observere direkte, der man i stedet bruker observerbare variabler for å estimere de uobserverbare konstruktene, gjerne kalt latente variabler. Wang og Wang (2012) peker på at det ikke bare er en utfordring at de latente variablene ikke kan undersøkes direkte, men at de observerte variablene ofte også inneholder betydelige målefeil. En av styrkene til SEM er at man kan ta høyde for disse målefeilene i modellene, slik at de latente variablene fremstår uten målefeil, men kun med sann varians (s. 1f). En annen stor fordel ved Mplus og SEM, er at man i tillegg, så sant modellen er overidentifisert, altså har et positivt antall frihetsgrader, kan få et estimat på hvor godt modellen passer til dataene.

I de konfirmerende faktoranalysene brukes rektordataene direkte, mens modellene som også inkluderer elevenes resultater fordrer at man først fletter elevens og rektorenes datasett, slikt at datasettet som benyttes inneholder både elevvariablene og rektorvariablene. Hver elev vil i dette nye datasettet ha individuelle data tilknyttet elevvariablene, men vil få lagt til svarene fra sin rektor på rektorvariablene. 30 elever fra samme skole vil følgelig ha individuelle svar på elevvariablene, men identiske svar på rektorvariablene. I forbindelse med modellene som undersøker sammenhengen mellom ledelseskonstruktene og elevenes resultater, vil standardfeilene som nevnt estimeres ved hjelp av replicate weights, mens det altså kun vil benyttes én plausible value for hver av elevenes resultater innen de tre fagområdene. Manglende data håndteres i de strukturelle modellene ved hjelp av Mplus7 sitt standardvalg, kalt FIML (*full information maximum likelihood*). Graham (2009) angir at FIML behandler de manglende dataene, utfører parameterestimerer og estimerer standardfeil i ett steg (s. 558) og anbefaler på det sterkeste at denne brukes hvis mulig (573).

3.2.2 Hva er CFA, og hvordan er dette blitt håndtert i studien?

Brown (2006) beskriver CFA som en form for SEM, som til forskjell fra eksplorerende faktoranalyser (EFA), som vil bli beskrevet senere, kjennetegnes av at alle aspekter ved analysen må angis på forhånd, basert på tidligere funn og teori. Dette dreier seg om hvilke faktorer man forventer å finne, hvilke indikatorer som forholder seg til disse faktorene og andre forhold som kan ha innvirkning på modellen. Belønningen kommer i form av et langt bredere analyseapparat enn det man får med EFA (s.1). Brown (2006) trekker som nevnt også frem at CFA er ”an indispensable analytic tool for construct validation” (s. 2), som også er hensikten ved å bruke CFA i denne studien. Det som gjør CFA så verdifullt, er muligheten for å ta høyde for målefeil og andre metodeeffekter når man estimerer hvor godt den teoretisk funderte modellen passer til de empiriske observasjonene (s. 2).

Mplus, herunder Mplus7, som er brukt i denne studien, er et av de mest brukte for gjennomføring av CFA, og er spesielt egnet for å håndtere blant annet mangelfulle datasett, ikke-normalitet og komplekse data (Wang og Wang, 2012, s. 24), som alle spiller inn i denne studien. Mplus tilbyr flere valgmuligheter i alle faser av analysen, og det er følgelig viktig at det redegjøres for de metodiske valgene som er tatt i nettopp denne studien:

- **Manglende data** er en utfordring i de fleste datasett, og det bør også i denne studien tas høyde for dette, da rundt 15% av datasettet tilknyttet rektorene savnes. Det nevnes ikke spesifikt hvilken metode PISA har brukt i sine SEM, men i SPSS-makroene brukes listwise deletion (OECD, 2009). I Mplus er som nevnt en nyere variant satt som standard for å håndtere dette, kalt FIML, og denne er også benyttet her.
- **Ikke-normalitet og små utvalg** preger deler av rektordatasettet, der utvalgsstørrelsen ligger på henholdsvis 196 og 207 for Norge og Sverige, som er i grenseland av hva som er ønskelig i faktoranalyser (Field, 2013, s. 684). I tillegg er normalfordelingskurvene til noen item sterkt forskjøvet. I møte med disse utfordringene kan og bør man bruke en såkalt robust estimator, og Wang og Wang (2012) trekker her frem MLR som et godt valg, da den reskalerer estimatene slik at de kan håndtere både ikke-normalitet og små utvalg (s. 60f). MLR brukes kun i forbindelse med CFA, da korrekte standardestimer i SEM, som nevnt, beregnes på bakgrunn av replicate weights.

3.2.3 Hva regnes som god modelltilpasning i denne studien?

Davey (2005) kan fortelle at modelltilpasning er et av de vanskeligste aspektene ved SEM, og at det eneste man i fagfeltet er enige om, er at man ikke kan se på én indikator for modelltilpasning alene, men at man må benytte seg av ulike kombinasjoner av disse såkalte fit-indikatorerne (s. 580). Målet med dette avsnittet er å beskrive fit-indikatorerne som brukes i denne studien, samt hva som anses som en god modelltilpasning i metodelitteraturen.

RMSEA *Root Mean Square Error of Approximation* er en badness-of-fit-indeks der tallet null angir den beste modelltilpasningen. Denne er i tillegg parsimoni-justert, noe som tilsier at den belønner et større antall av frihetsgrader, som er et kjennetegn ved mindre komplekse modeller. (Wang og Wang, 2012, s. 19).

SRMR *Standardized Root Mean square Residual* angir den gjennomsnittlige kvadrerte avstanden mellom den observerte og den modellestimerte korrelasjonsmatrisen (Wang og Wang, 2012, s. 20). Verdien null angir perfekt modelltilpasning.

CFI *Comparative Fit Index* sammenligner den spesifiserte modellen med null-modellen, som antar at det ikke er noen kovarians mellom de observerte variablene. CFI angis fra 0 til 1. Jo nærmere 1, jo bedre tilpasning (Wang og Wang, 2012, s. 19).

TLI *Tucker-Lewis Index* angis som CFI fra 0 til 1, og er også en komparativ fit-indeks. Forskjellen er at kji-kvadratet og frihetsgradene spiller inn, slik at den, i likhet med RMSEA, vil belønne mindre komplekse modeller (Wang og Wang, 2012, s. 19).

Hu og Bentler (1999) angir grenseverdier for hver av de nevnte indeksene, i tillegg til at en rekke kombinasjoner er undersøkt. Det finnes kritikere som mener at disse grenseverdiene ikke må overgeneraliseres (Marsh, Hau, og Wen, 2004), men bruken av disse er fortsatt vanlig praksis (Brown, 2006; Byrne, 2012; Kline, 2011; Thompson, 2004). Grensene som angis hos Hu og Bentler (1999) er $RMSEA \leq .06$, $SRMR \leq .08$ og $CFI/TLI \geq .95$. Kombinasjoner av disse øker påliteligheten, og spesielt fremheves $SRMR \leq .08$ i sammenheng med enten $CFI/TLI \geq .95-.96$ og/eller $RMSEA \leq .06$, selv om disse i noen få tilfeller kan forkaste også gode modeller når $N \leq 250$ (s. 24ff). Disse grenseverdiene vil brukes også her, men det bør i den forbindelse nevnes at kun SRMR og RMSEA er tilgjengelige ved bruk av RW. I tillegg til disse fire, vil kji-kvadratstester benyttes når det er nødvendig å undersøke om endringer i tilpasningen mellom modeller med ulikt antall frihetsgrader er signifikante.

3.3 Analyser utført med SPSS 21

IBMs statistikkprogram SPSS er det mest brukte statistikkprogrammet på markedet, men kan altså ikke brukes til strukturell modellering. Derimot har det flere andre nyttige bruksområder, og i tillegg til generell behandling av datafilene fra PISA og enkle frekvensanalyser av rektorenes svar har SPSS blitt brukt i to andre sammenhenger i denne studien.

3.3.1 Hva er reliabilitet, og hvordan er dette håndtert i studien?

Det anbefales at man alltid rapporterer reliabiliteten i forbindelse med indre konsistens (Kline, 2011, s. 70), og den mest brukte algoritmen for reliabilitet er Cronbachs Alpha (α). Denne angir på en skala fra 0 til 1 hvor konsekvente respondentene er når de svarer på item som tilhører samme konstrukt. Cohen et al. (2011) angir .7 som grense for akseptabel reliabilitet, .8 for høy reliabilitet og .9 for meget høy reliabilitet (s. 640). Hallinger (1994) oppgir .8 som en anbefalt grense i forbindelse med konstruktene i PIMRS (s. 6), mens Kjærnsli og Olsen (2013b) opererer med det noe vagere ”rundt eller over .7” som en nedre grense i den norske PISA-rapporten fra PISA 2012. De presiserer videre at en reliabilitet på .7 indikerer at 70% av variansen er ”sann varians”, mens 30% er ”feilvarians”, som inkluderer både målefeil og ekte, unik varians (s. 316f). Selv om denne studien fokuserer mer på modelltilpasningen som helhet enn den indre konsistensen til hvert av konstruktene, er det ingenting i veien for å rapportere og vurdere Cronbachs α for konstruktene, da dette også gjøres i rapportene fra PISA 2012.

3.3.2 Hva er EFA, og hvordan er dette blitt håndtert i studien?

Eksplorerende faktoranalyser (EFA) skiller seg fra CFA ved at sistnevnte altså er teoridrevet, og sammenligner en teoretisk fundert datamatrix med den observerte datamatriksen. EFA er derimot datadrevet, og man trenger derfor heller ikke ha noen forventning om eventuelle konstrukter, såkalte faktorer. Selv om man eventuelt skulle ha slike forventninger, kan man gjennomføre analysen uten å ta hensyn til disse (Thompson, 2004, s. 6), noe som kan sies å gjelde i denne studien. Field (2013) forklarer at faktoranalysen leter etter grupper av variabler som korrelerer høyt med hverandre, men som korrelerer langt mindre med de andre variablene (s. 668), og SPSS stiller opp de viktigste faktorene i datasettet og viser samtidig hvordan itemene lader opp mot disse. Målet med en slik analyse er å lete etter såkalt multidimensjonalitet og således redusere antall faktorer i datasettet.

På tross av at man ikke trenger å ha noen forventninger om eventuelle faktorer på forhånd, er det likevel noen metodiske valg man må ta stilling til før analysene gjennomføres:

- **PAF eller PCA:** Det finnes flere varianter av EFA. De to vanligste er *Principal Axis Factoring* (PAF) og *Principal Components Analysis* (PCA), og i denne studien har det vært mest naturlig å bruke PAF, da denne utelukker den unike variansen fra analysen (Field, 2013, s. 675). Dette kan være en fordel fordi det kan bidra til å redusere målefeil (Bryman og Cramer, 2011, s. 322). Costello og Osborne (2005) konkluderer også med at PAF er mest egnet for samfunnsvitenskapene, spesielt ved ikke-normalitet (s. 2).
- **Manglende data** er, med bakgrunn i det lave antallet rektorer i denne studien, håndtert ved å benytte *pairwise deletion*, som bruker alle tilgjengelige målepunkter, i stedet for å se bort fra alle caser der ett eller flere målepunkter mangler, som er tilfellet med *listwise deletion*. Graham (2009) peker på at begge disse kan være problematiske, men konkluderer med at *pairwise deletion* godt kan benyttes i en utforskende fase (s. 554).
- **Faktorrotasjon** har som mål å tydeliggjøre forskjellene mellom faktorene og forenkle fortolkningsprosessen. Field (2013) legger vekt på at *oblique rotasjon* gir flere muligheter for å tilpasse faktorrotasjonen, fordi den tillater at faktorene kan korrelere (s. 680). I de fleste tilfeller vil man i samfunnsvitenskapene også forvente at faktorene er korrelert, og derfor anbefales også ofte *oblique rotasjon* (Bryman og Cramer, 2011; Costello og Osborne, 2005; Field, 2013). I SPSS brukes varianten kalt *oblique oblmin*, og i denne studien rapporteres den tilhørende *pattern matrix* fra den roterte faktoranalysen.
- **Antall faktorer som skal beholdes** er også noe som kan by på utfordringer. En vanlig regel å følge her, er det såkalte Kaiser-kriteriet, hvor man beholder alle faktorer som har en *eigenvalue* på større enn 1 (Costello og Osborne, 2005, s. 1). Field (2013) forklarer at en faktor med *eigenvalue* over 1 har en positiv andel av systematisk over usystematisk varians, og SPSS gir oss *eigenvalue*ne til faktorene, samt den prosentvise andelen av den totale variansen som forklares av faktorene. Ettersom denne studien fokuserer på både psykometriske analyser og teoretiske drøftinger tilknyttet faktorene, beholdes alle faktorer med *eigenvalue* over 1 som representerer et relevant aspekt ved pedagogisk ledelse, selv om Costello og Osborne (2005) viser til at bruk av Kaiser-kriteriet alene kan føre til at man i noen tilfeller tar med for mange faktorer videre i analysen (s. 2).

3.4 Forskningsetikk

Cohen et al. (2011) er klare på at alle større studier innen utdanningsforskning bærer i seg utfordringer tilknyttet forskningsetikk, men deres omfattende liste over mulige etiske implikasjoner (s. 165), er i liten grad relevante med tanke på denne studien. Dette skyldes at det ikke har foregått noen aktiv innhenting av data, og at datasettene som er blitt brukt er åpent tilgjengelige og anonymisert. Det har således vært umulig å få tilgang til og følgelig publisere noe som kan spores tilbake til og få konsekvenser for de som har deltatt i studien.

Neste skritt blir derfor å vurdere hvorvidt PISA-undersøkelsene forholder seg til forskningsetiske retningslinjer som er aktuelle for denne studien. Det finnes i utgangspunktet ingen omtale av dette i noe av rammeverket eller rapportene tilknyttet PISA 2012, men OECD (2011) gir oss et innblikk i prinsippene som ligger til grunn for OECD sitt arbeid innen statistikk. OECD hevder selv å ha ”a high reputation for quality og integrity around the world” (s. 3), og OECDs hovedprinsipper skal bidra til å sikre dette. Av det som er aktuelt i forbindelse med denne studien, er det viktig at statistikken skal innhentes i tråd med ”professional ethics with regard to methods and procedures used for collecting, processing, storage and dissemination of statistical data”. Videre kan man lese at ”individual data collected by the OECD for statistical compilation are considered to be strictly confidential”, og at “specific measures are taken to ensure the full protection of confidential data from any potential disclosure”. Sist, men ikke minst er OECD “committed to carrying out its statistical activities in co-ordination with national statistical agencies” (s. 11).

Det går ikke ytterligere i detalj på prosedyrene som brukes i de ulike fasene av datainnsamlingsfasen for å forsikre at disse prinsippene følges, men den tekniske rapporten til PISA 2009 tyder på at prinsippene er vel ivaretatt (OECD, 2012). Dette gjelder også nasjonale hensyn som innhenting av nødvendige tillatelser fra NSD og elevenes foresatte. Sivesind og Elstad (2010) er ett av mange eksempler på at det stadig settes et kritisk søkelys på PISA-undersøkelsene, men ut over at elevene kan bli stresset av testsituasjoner på generell basis, er det svært lite kritikk å finne her eller andre steder som dreier seg om forskningsetiske overtredelser. Denne mangelen på kritikk indikerer at det finnes en bred forståelse for at PISA gjennomføres i tråd med hovedprinsippene for OECD sitt statistiske arbeid.

4 RESULTATER

Resultatkapittelet vil gjenspeile strukturen fra forskningsspørsmålene, og hvert av delkapitlene nedenfor innledes med forskningsspørsmålet som skal belyses, før de tilhørende resultatene som er nødvendige for senere å diskutere forskningsspørsmålet vil bli presentert og gjennomgått. Selv om resultatkapittelet fokuserer på de kvantitative resultatene, vil det likevel være nødvendig med enkelte innholdsmessige analyser underveis, da noen av de statistiske analysene gjøres på bakgrunn av substansiell vurderinger. Første delkapittel vil inneholde psykometriske analyser tilknyttet reliabilitet og modelltilpasning for konstruktene i PISA 2012 sin Index of School Management. Neste del vil fokusere på etableringen av alternative ledelseskonstrukt i PISA 2012. I tråd med den tidligere presentasjonen av dette, dreier det seg her om eksplorerende faktoranalyser, etterfulgt av konfirmerende faktoranalyser i tillegg til en reliabilitetsanalyse av de nye konstruktene. Kapittelets tredje del vil inneholde frekvensanalyser av itemene som inngår i konstruktene, som primært skal bidra til å belyse likheter og forskjeller mellom svarene til de norske og de svenske rektorene. Fjerde og siste del av resultatkapittelet vil presentere de strukturelle modellene som undersøker sammenhengen mellom rektors pedagogiske ledelse og elevenes resultater.

4.1 Analyse av PISA-konstruktene

→ Har konstruktene som inngår i PISA 2012 sin Index of Principal Management en tilfredsstillende begrepsvaliditet sett i lys av svarene til norske og svenske rektorer?

PISA opererer med fire konstrukt basert på 13 av i alt 21 item om pedagogisk ledelse i skole spørreskjemaet (SC34). Disse konstruktene er *Framing and communicating the school's goals and curricular development* (LEADCOM), *Instructional leadership* (LEADINST), *Promoting instructional improvements and professional development* (LEADPD), og *Teacher participation* (LEADTCH) (OECD, 2013b, s. 140). I denne studien brukes Chronbachs α til å angi den indre konsistensen til konstruktene, men da dette ikke alltid er tilstrekkelig for å konkludere med at konstruktene passer med de empiriske observasjonene, i dette tilfellet rektorenes svar, vil konstruktene også analyseres gjennom en konfirmerende faktoranalyse.

4.1.1 Hvordan var reliabiliteten til PISA-konstruktene?

Tabell 4-2 nedenfor lister opp de fire PISA-konstruktene og deres tilhørende item. SPSS er her brukt for å kalkulere Cronbachs α for de fire konstruktene, basert på rektordataene.

LEAD_	ITEM	α	
		NOR	SVE
COM	SC34Q02: Jeg bruker elevenes faglige resultater for å utvikle skolens pedagogiske mål.	.796	.780
	SC34Q03: Jeg forsikrer meg om at lærernes utvikling er i samsvar med skolens pedagogiske mål.		
	SC34Q04: Jeg sørger for at lærernes arbeid er i samsvar med skolens pedagogiske mål.		
	SC34Q14: Jeg diskuterer skolens faglige mål med lærere på fellesmøter.		
INST	SC34Q05: Jeg oppfordrer til undervisningspraksis basert på nyere forskning.	.772	.738
	SC34Q06: Jeg roser lærere som har elever som deltar aktivt i undervisningen.		
	SC34Q08: Jeg gjør lærere oppmerksomme på viktigheten av å utvikle elevenes kritiske og sosiale evner.		
PD	SC34Q07: Jeg tar initiativ til å diskutere årsakene når en lærer har problemer i klasserommet sitt.	.777	.801
	SC34Q09: Jeg er opptatt med forstyrrende adferd i klasserommet.		
	SC34Q13: Når en lærer kommer med et problem knyttet til arbeidet i klasserommet, løser vi det sammen.		
TCH	SC34Q10: Jeg gir ansatte muligheten til å delta i beslutningsprosesser angående skolen.	.709	.689
	SC34Q11: Jeg motiverer lærere til å utvikle en skolekultur basert på stadige forbedringer.		
	SC34Q12: Jeg ber lærere delta i evalueringen av ledelsens arbeid.		

Tabell 4-2: Konstruktene i PISA sin Index of School Management med tilhørende verdier for Cronbachs α

Alfa-koeffisienten ligger i alle bortsett fra én over den mye brukte grenseverdien på .7, selv om Kjærnsli og Olsen (2013b) opererer med ”rundt eller over .7”. Med et lite unntak av LEADTCH for Sveriges del, er det altså anerkjent praksis å gå videre med disse konstruktene til neste ledd i begrepsvalideringen, nemlig en innholdsanalyse. I denne studien er det derimot lagt vekt på at konfirmerende faktoranalyser kan gi et mer fullstendig bilde av hvor godt konstruktene, og den teoretisk baserte modellen, samsvarer med de empiriske observasjonene.

4.1.2 Hvordan ble modelltilpasningen med PISA-konstruktene?

De fire fit-indikatorerne og de tilhørende verdiene fra den konfirmerende faktoranalysen er fremstilt i Tabell 4-3 under. Analysene er gjennomført i tråd med retningslinjene i metoddelen, og modellene som er testet tilsvarende som vises i Figur 4-2 og 4-3 lenger ned.

	RMSEA	SRMR	CFI	TLI
NORGE	.089	.067	.912	.884
SVERIGE	.080	.062	.907	.877

Tabell 4-3: Modelltilpasning for CFA med PISA-konstruktene

Kun én av fit-indeksene, markert med grønt, holder seg innenfor retningslinjene for god modelltilpasning hos Hu og Bentler (1999), som tilsier at RMSEA og SRMR bør ligge så nærme 0 som mulig, der henholdsvis .08 og .06 utgjør en øvre grense, mens CFI/TLI bør

ligge så nærme 1 som mulig, med .95 som nedre grense (s. 24ff). Det er således ikke anbefalt å gå videre med strukturelle modeller basert på disse konstruktene, på tross av at konstruktene med kun ett unntak hadde akseptabel reliabilitet. De konfirmerende faktoranalysene som gjengis i tabell 4-3 ble for ordens skyld også gjennomført med flettede datafiler, der rektorenes svar analyseres i kraft av elevene de representerer. Når elevene fikk tildelt variablene fra deres respektive rektorer, og disse så ble analysert, viste RMSEA en svakere modelltilpasning for begge land. For Norges del gikk RMSEA fra .089 til .115, og for Sveriges del fra .080 til .114. SRMR forble uendret, for Norges del, men ble for Sveriges del marginalt svakere, med .064, mot tidligere .062. CFI og TLI er som nevnt ikke tilgjengelige ved bruk av replicate weights i Mplus7. Årsakene til denne endringen har det ikke vært mulig å komme helt til bunns i, men det kan skyldes at utvalgsstørrelsene går kraftig opp, at det er noen aspekter ved tuestegssamplingen som rektordataene alene ikke tar høyde for, eller det kan ha en sammenheng med at alle elevene på samme skole får samme verdi på en rekke item, som sammen med bruken av replicate weights, kan få en slik effekt. Dette fenomenet er uansett ikke av stor betydning her, da modelltilpasningen i begge tilfeller er uakseptabel.

Avslutningsvis kan det også nevnes at kun LEADCOM har tilstrekkelig med item tilknyttet konstrukt til at det er overidentifisert, og følgelig kan testes enkeltstående med en CFA. Resultatet for dette konstruktet med norske data ble RMSEA = .108, SRMR = .025, CFI = .978 og TLI = .933. SRMR og CFI er her akseptable, men ingen av de to parsimoni-justerte indeksene holder seg innenfor grenseverdiene satt av Hu og Bentler (1999), noe som medfører at ingen av de anbefalte kombinasjonene av fit-indikatorer oppfylles med disse verdiene. Modelltilpasningen ble svært lik med svenske data, med samme konklusjon.

4.2 Etablering av alternative konstrukt

→ Kan man, basert på itemene tilknyttet pedagogisk ledelse i PISA 2012, etablere alternative ledelseskonstrukt med en mer tilfredsstillende begrepsvaliditet?

Dette delkapittelet innledes av eksplorerende faktorundersøkelser med norske og svenske data. Så vil analysene som ledet frem mot et felles sett med alternative konstrukt gjennomgå, og det vil fortløpende bli redegjort for de teoretiske vurderingene som ble gjort underveis. Når de alternative konstruktene er etablert, vil det foretas en reliabilitetsanalyse av konstruktene, i tillegg til at de konfirmerende faktoranalysene gjennomgås nærmere.

4.2.1 Hva ble resultatet av de eksplorerende faktoranalysene?

I forkant av analysene ble det anslått hvor godt dataene egnede seg for faktoranalyser gjennom en *Keiser-Meyer Olkin Measure of sampling adequacy*, der verdiene for de norske og svenske datasettene var på henholdsvis .898 og .887, noe som anses som svært godt (Field, 2013, s. 685). De eksplorerende faktoranalysene ble deretter gjennomført i tråd med de metodiske vurderingene som ble presentert i metodedelen over. De eksplorerende faktoranalysen tok utgangspunkt i alle de 21 itemene tilknyttet rektors pedagogiske ledelse, men item som ladet med tilsvarende styrke på flere faktorer samt item som ikke ladet med over .3 på noen faktor, ble utelatt én etter én, før analysene ble kjørt på nytt. Tabell 4-4 og 4-5 viser den endelige løsningen for hvert land, og faktorene stilles opp etter eigenvalue, med itemene sortert innad i hver faktor etter hvor sterkt de lader mot faktoren. Ordlyden i itemene skrives av plasshensyn ikke fullt ut, men itemene er gjengitt i sin helhet i Vedlegg 7.1, der også itemene som falt utenfor i faktoranalysen kan finnes.

NORSKE REKTORER

Pattern Matrix fra eksplorerende faktoranalyse (PAF) - NORGE		Faktor			
		1	2	3	4
SC34Q02 – Elevenes resultater > Skolens mål		.835			
SC34Q03 – Lærernes utvikling > < Skolens mål		.762			
SC34Q04 – Lærernes arbeid > < Skolens mål		.716			
SC34Q05 – Undervisningspraksis i tråd med nyere forskning		.522			
SC34Q01 – Fremme skolens omdømme i lokalsamfunnet.		.415			
SC34Q13 – Lærer kommer med klasseromsproblem som løses sammen			.817		
SC34Q07 – Tar initiativ til å diskutere problemer i klasserommet			.708		
SC34Q09 – Opptatt med problemer med forstyrrende klasseromsadferd			.590		
SC34Q20 – Vurderer elevarbeider når undervisningen evalueres				.631	
SC34Q19 – Jevnlige, uformelle observasjoner i klasserommene				.560	
SC34Q22 – Vurderer lærernes innsats				.559	
SC34Q12 – Ber lærerne delta i evalueringen av ledelsens arbeid					-.609
SC34Q15 – Henviser til lærepl. når lærere deltar i lærepl.avgjørrelser					-.583
SC34Q16 – Diskuterer læringsresultater med lærere når faglaner ...					-.564
Eigenvalue		5.703	1.236	1.161	1.047
Prosentandel forklart varians		40.73	8.83	8.30	7.48
Extraction Method: Principal Axis Factoring.					
Rotation Method: Oblimin with Kaiser Normalization (Rotation converged in 7 iterations).					
Ladninger på under .3 er utelatt					

Tabell 4-4: Eksplorerende faktoranalyse - Norske rektorer

Den norske faktorløsningen, som presenteres i Tabell 4-4, endte opp med å inkludere 14 av 21 item. De fire fargekodene viser hvilke item som også kan finnes i den svenske faktorløsningen i Tabell 4-5, mens hvite item følgerlig kun finnes i den norske løsningen.

Nederst i tabellen vises faktorenes eigenvalue, samt prosentandelen av forklart varians. Den eksplorerende faktoranalysen av de norske rektordataene endte opp med fire faktorer med eigenvalue over 1. De fire faktorene forklarer til sammen 68,2% av den totale variansen.

SVENSKE REKTORER

Pattern Matrix fra eksplorerende faktoranalyse (PAF) - SVERIGE		Faktor			
		1	2	3	4
SC34Q04 – Lærernes arbeid > < Skolens mål		.932			
SC34Q05 – Undervisningspraksis i tråd med nyere forskning		.765			
SC34Q03 – Lærernes utvikling > < Skolens mål		.615			
SC34Q02 – Elevenes resultater > Skolens mål		.607			
SC34Q11 – Motiverer til kultur basert på stadige forbedringer.		.386			
SC34Q19 – Jevnlige, uformelle observasjoner i klasserommene			.895		
SC34Q20 – Vurderer elevarbeider når undervisningen evalueres			.722		
SC34Q22 – Vurderer lærernes innsats			.367		
SC34Q07 – Tar initiativ til å diskutere problemer i klasserommet				.858	
SC34Q13 – Lærer kommer med klasseromsproblem som løses sammen				.732	
SC34Q09 – Opptatt med problemer med forstyrrende klasseromsadferd				.661	
SC34Q17 – Leder eller deltar i etterutdanningsaktivitet med lærerne					.796
SC34Q18 – Setter av tid til erfaringsutveksling					.747
SC34Q14 – Diskuterer skolens faglige mål med lærerne på fellesmøter					.479
SC34Q16 – Diskuterer læringsresultater med lærere når faglaner					.417
SC34Q15 – Henviser til lærepl. når lærere deltar i lærepl.avgjørrelser					.378
Eigenvalue		6.722	1.524	1.500	1.184
Prosentandel forklart varians		39.54	8.97	8.82	6.96
Extraction Method: Principal Axis Factoring.					
Rotation Method: Oblimin with Kaiser Normalization (Rotation converged in 6 iterations.)					
Ladninger på under .3 er utelatt					

Tabell 4-5: Eksplorerende faktoranalyse - Svenske rektorer

I den svenske faktorløsningen, presentert over i Tabell 4-5, var det totalt 5 item som ikke ble en del av den endelige faktorløsningen. Fargekodene angir, som for Tabell 4-4, hvilke item som inngår for begge land, mens hvite item viser hvilke item som kun er en del av den svenske faktorløsningen. Resultatet ble også her fire faktorer, dog med en noe ulik fordeling av item. Tabellen viser et tilsvarende bilde med tanke på eigenvalue og forklart varians som den norske faktorløsningen, og de fire faktorene forklarer her 64,3% av den totale variansen.

4.2.2 Hvilke item inngår i de alternative konstruktene?

I denne delen presenteres prosessen som ledet fra de eksplorerende faktoranalysene til de alternative konstruktene. I de tilfellene der konkurrerende utgaver av konstruktene skulle testes mot hverandre, ble dette gjort gjennom å sammenligne modelltilpasningene og

kjikkvadratene fra konfirmerende faktoranalyser av de ulike variantene, samt ved å analysere reliabiliteten til ulike utgaver av konstruktet. Noen steder spilte også substansielle vurderinger inn, noe som innebærer at innholdsdimensjonen av begrepsvalideringen blir berørt allerede her. Teoretiske analyser som er av betydning for konstruktens validitet, men som ikke var relevante i denne prosessen, spares derimot til diskusjonsdelen. Faktorene nummereres i det følgende i tråd med den norske tabellen, Tabell 4-4, som kun har en betydning for faktor 2 og 3, som endte opp i omvendt rekkefølge i den svenske faktorløsningen, presentert i Tabell 4-5.

FAKTOR 1

Denne faktoren inneholder fire item som er felles både i Norge og Sverige, SC34Q02-SC34Q05. De tre første av disse har en tydelig tilknytning til skolens mål, og inngår også i LEADCOM. SC34Q05 skiller seg derimot ut ved at skolens mål ikke er eksplisitt nevnt. Det faktum at dette itemet statistisk sett henger godt sammen med de tre andre, kan tyde på at ”undervisning basert på nyere forskning” mange steder inngår som ett av skolens mål, og at det å oppfordre til dette kan anses som en del av det å arbeide med skolens mål. En utelatelse av SC34Q05 fører også til at Cronbachs α blir marginalt svakere for begge land. For Norges del dreier det seg om en endring fra .862 til .858, mens reliabiliteten for Sveriges del går fra .842 til .820. På bakgrunn av dette ble itemet tatt med i konstruktet inntil videre. Et tilsvarende argument kunne også bli brukt om SC34Q11, som inngår i den svenske faktoren, men da itemet viste seg å svekke reliabiliteten til konstruktet, ble det ikke tatt med. SC34Q01, som bare inngår i den norske faktorløsningen, viser i tillegg en svak tematisk forbindelse til de andre, og ble følgelig også utelatt. Dette første alternative konstruktet fikk navnet MÅL.

FAKTOR 2

Både den norske og den svenske faktorløsningen har med denne faktoren, og alle tre itemene inkluderes følgelig i konstruktet. Faktoren er i tillegg identisk med PISA-konstruktet LEADPD, men av to årsaker har konstruktet likevel fått et annet navn. Det viktigste argumentet for dette er at itemene kan sies å omhandle utfordringer med læringsmiljøet i klasserommet, vel så mye som de peker mot lærerens undervisning og profesjonelle utvikling, selv om disse ikke er uavhengige av hverandre. Dette vil bli diskutert ytterligere i den innholdsmessige delen av begrepsvalideringen. For det andre kan det av praktiske årsaker være hensiktsmessig å tydelig skille PISA-konstruktene fra de alternative konstruktene. Dette konstruktet utgjør altså ikke et alternativ i seg selv, da itemene det består av ikke er endret i

forhold til PISA-konstruktet, men det er like fullt en del av konstruktene som sammen danner et alternativ. Dette andre konstruktet ble gitt navnet LÆRINGSMILJØ.

FAKTOR 3

Den tredje faktoren inneholder tre item som ikke finnes i noen av PISA-konstruktene. Ettersom alle itemene inngår i samme faktor i begge land, og alle er relatert til ulike aspekter ved rektors evaluering av skolens praksis, ble faktoren tatt med videre, og konstruktet fikk navnet EVALUERING.

FAKTOR 4

Proessen frem mot det siste alternative konstruktet var mer kompleks enn for de foregående, da de to faktorløsningene her var mindre samstemte. Den første avgjørelsen som ble tatt, var at SC34Q17 og SC34Q18 skulle utelukkes, da de kun inngår i den svenske faktoren, samt at de ikke omhandler situasjoner der lærerne medvirker i evaluering- eller beslutningsprosesser ved skolen, som ellers var et gjennomgående tema. Både SC34Q15 og SC34Q16 hadde derimot en tilknytning til dette medvirkningsperspektivet, og var i tillegg å finne i begge faktorløsningene. SC34Q12 og SC34Q14 har det til felles at de begge omhandler lærermedvirkning, men kun er å finne i én av faktorløsningene, henholdsvis den norske og den svenske. Ettersom disse begge kan være av interesse, kunne konstruktet potensielt sett inneholde opp til fire konstrukt i ulike kombinasjoner. Det ble derfor besluttet å kjøre konfirmerende faktoranalyser, igjen tilsvarende Figur 4-2 og 4-3, med alle de potensielt interessante kombinasjonene av item og resultatene av disse analysene er gjengitt i Tabell 4-6. Ettersom kombinasjonen av RMSEA og SRMR i kombinasjon er velegnet for å gi et godt inntrykk av modelltilpasningen, er CFI og TLI for enkelhets skyld utelatt i tabellen nedenfor.

ITEM	NORGE		SVERIGE	
	RMSEA	SRMR	RMSEA	SRMR
12 + 14 + 15 + 16	.046	.046	.053	.055
12 + 14 + 15	.044	.044	.053	.054
12 + 14 + 16	.053	.047	.051	.052
12 + 15 + 16	.026	.037	.043	.053
14 + 15 + 16	.050	.046	.055	.054
15 + 16	.030	.036	.045	.053
Uten konstruktet	.032	.036	.044	.049

Tabell 4-6: Modelltilpasning med de alternative faktorsammensetningene for Faktor 4

Tabell 4-6 viser at kombinasjonen av SC34Q12, SC34Q15 og SC34Q16, markert med grønt, ser ut til å være den mest velegnede kombinasjonen, i det minste sammenlignet med de andre modellene med tre item i konstruktet. Når antall frihetsgrader i modellen endres, som gjelder

kombinasjonene markert med gult, er det derimot ikke nødvendigvis nok å vurdere fit-indikatorerne alene, selv om RMSEA tar høyde for dette ved å favorisere modeller med et høyere antall frihetsgrader. For å være på den sikre siden, kan modeller med ulikt antall frihetsgrader testes mot hverandre ved hjelp av kjikvadratstester. Disse må regnes ut manuelt, da Mplus7 ikke kan utføre slike analyser når man bruker MLR-estimatoren³. Her vil en p-verdi på under .05 bety at den mindre restriktive modellen er signifikant bedre (på .05-nivå) enn den med et høyere antall frihetsgrader, mens en p-verdi på over .95 indikerer at den mer restriktive modellen er signifikant bedre (på 0.5-nivå) enn den med færre frihetsgrader.

Når modellen som inkluderer konstruktet markert i grønt testes mot modellen der konstruktet utelates fullstendig, gir dette p-verdier på henholdsvis .415 og .096 for Norge og Sverige, som indikerer at den sistnevnte modellen ikke er signifikant bedre. Testes den samme modellen mot den med fire item i konstruktet, ser vi derimot at den mindre restriktive modellen med tre item er signifikant bedre enn den med fire item, med p-verdier på .001 og .003. Det siste alternativet, modellen med to item i konstruktet, er heller ikke signifikant bedre, med p-verdier på .486 og .121. Det å inkludere SC34Q12 i konstruktet, sammen med SC34Q15 og SC34Q16, er med andre ord forsvarlig, selv basert på svenske data, der itemet ikke var en del av faktorløsningen etter den eksplorerende faktoranalysen. Konklusjonen blir derfor at konstruktet markert med grønt videreføres, da en modell basert på dette konstruktet ikke er signifikant svakere enn de andre modellene, og i tillegg beskriver en relevant side ved rektors pedagogiske ledelse. Dette siste konstruktet fikk navnet MEDVIRKNING.

ET NYTT BLIKK PÅ SC34Q05

Nå som det fjerde alternative konstruktet også er etablert, kan inkluderingen av SC34Q05 i MÅL også undersøkes ved hjelp av konfirmerende faktoranalyser og kjikvadratstester. Når itemet fjernes, går RMSEA for Norges del fra .026 til .042, mens man for Sveriges del får en motsatt, men svakere effekt, der RMSEA går fra .042 til .035. Modellene holder seg altså trygt innenfor grenseverdiene til Hu og Bentler (1999) også etter utelatelsen. Kjikvadratstesten viser for Norges del at modellen blir signifikant svakere ved å utelate SC34Q05, med en p-verdi på .995, som er signifikant på .01-nivå. For Sveriges del blir det også her omvendt, med en p-verdi på .027, som betyr at modellen med færre frihetsgrader er signifikant bedre, riktignok bare på .05-nivå. Utelatelsen av SC34Q05 medfører altså en signifikant endring for begge land, dog med motsatt fortegn. Ettersom signifikansen som taler for å inkludere itemet

³ Se <http://www.statmodel.com/chidiff.shtml> for en detaljert presentasjon av utregningsmetoden.

vår på .01-nivå, og endringen i RMSEA også var større i den norske modellen, videreføres konstruktet som det var, inkludert SC34Q05. Det at modellen primært skal leses i norsk kontekst, kan også brukes som et argument for en slik avgjørelse.

4.2.3 Hva er konstruktene psykometriske egenskaper?

I tabellene under ser vi resultatet av en reliabilitetsanalyse av de alternative konstruktene. Konstruktene og de tilhørende itemene er fremstilt på samme måte som for PISA-konstruktene over, og baserer seg også her på en analyse av rektordataene.

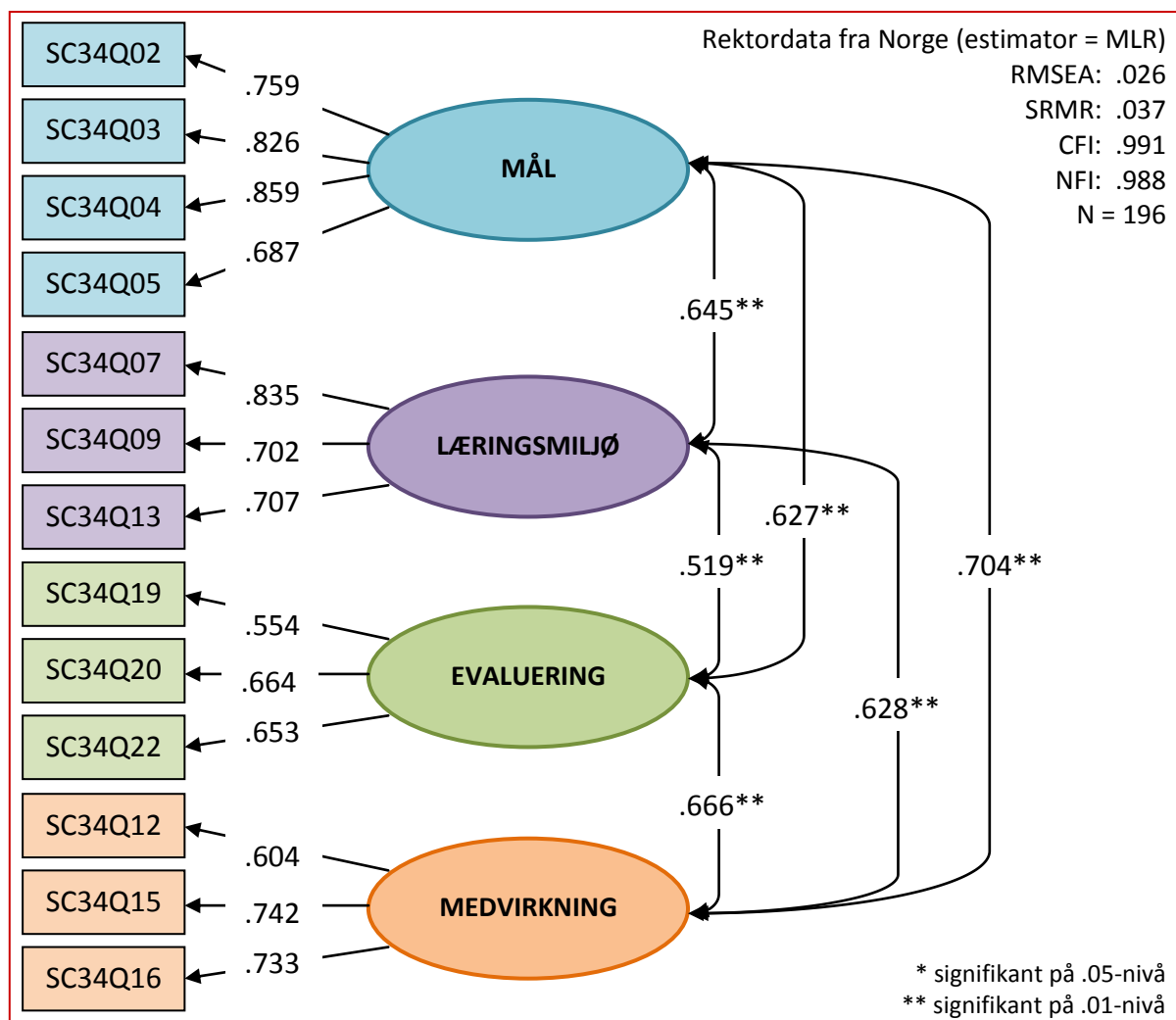
	ITEM	α NOR SVE
MÅL	SC34Q02: Jeg bruker elevenes faglige resultater for å utvikle skolens pedagogiske mål.	.862
	SC34Q03: Jeg forsikrer meg om at lærernes utvikling er i samsvar med skolens pedagogiske mål.	
	SC34Q04: Jeg sørger for at lærernes arbeid er i samsvar med skolens pedagogiske mål.	.842
	SC34Q05: Jeg oppfordrer til undervisning basert på nyere forskning.	
LÆR	SC34Q07: Jeg tar initiativ til å diskutere årsakene når en lærer har problemer i klasserommet sitt.	.777
	SC34Q09: Jeg er opptatt med forstyrrende adferd i klasserommet.	.801
	SC34Q13: Når en lærer kommer med et problem knyttet til arbeidet i klasserommet, løser vi det sammen.	
EVAL	SC34Q19: Jeg gjennomfører uformelle observasjoner i klasserommet jevnlig,	.654
	SC34Q20: Jeg vurderer elevarbeider når jeg evaluerer undervisningen.	.716
	SC34Q22: Jeg vurderer lærernes innsats.	
MEDV	SC34Q12: Jeg ber lærere delta i evalueringen av ledelsens arbeid.	.736
	SC34Q15: Jeg henviser til skolens faglige mål når lærere deltar i avgjørelser angående fagplaner.	.701
	SC34Q16: Jeg diskuterer læringsres. med lærerne for å kartlegge styrker og svakheter i skolens fagplaner.	

Tabell 4-7: Oversikt over konstrukt i PISA med tilhørende item og Cronbachs α

Det alternative konstruktet EVALUERING er det eneste som faller under .7, dog kun for Norges del. Reliabiliteten er med andre ord lavere enn det som ofte brukes som en nedre grense, selv om den kan sies å tangere mot Kjærnsli og Olsen (2013b) sin grenseverdi ”rundt eller over .7”. På bakgrunn av dette ble det også her utført en kjikvadratstest, der modellen basert på alle de fire konstruktene ble testet mot en modell der EVALUERING ble utelatt. Med norske og svenske data ble p-verdiene på henholdsvis .092 og .074, som betyr at utelatelsen strengt tatt ikke gir en signifikant forbedring. Ettersom konstruktet også representerer et aspekt ved rektors ledelse som kan være av interesse, videreføres konstruktet på tross av den svake reliabiliteten.

Nedenfor fremstilles de konfirmerende faktoranalysene med standardiserte ladninger og korrelasjoner. Av plassmessige hensyn er den unike variansen til itemene utelatt fra modellen, men denne kan lett beregnes ved å kvadrere ladningen mot konstruktet og trekke dette fra 1.

NORSKE REKTORER

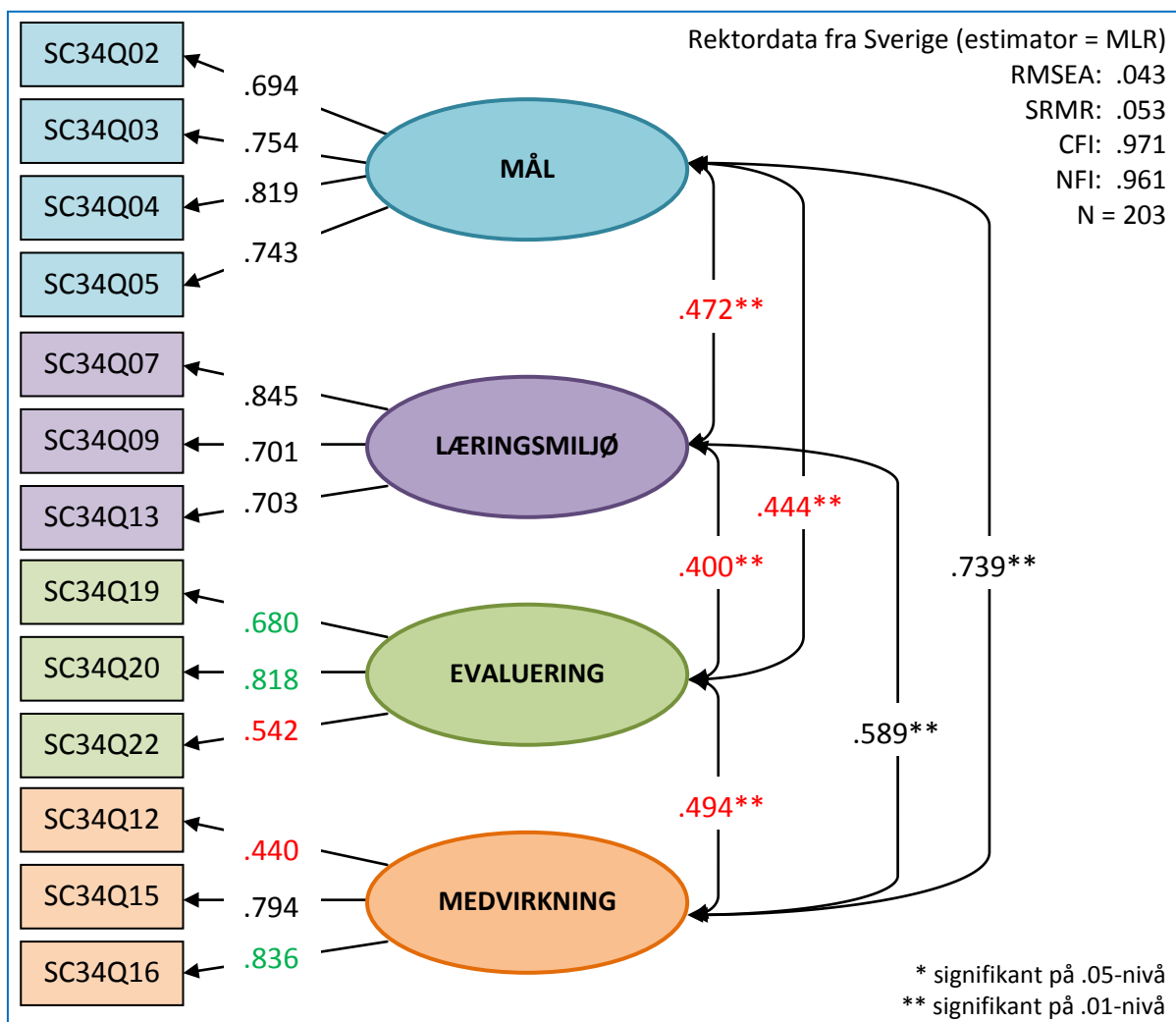


Figur 4-2: Konfirmerende faktoranalyse av alternative konstrukturer med rektordata - Norge

Modelltilpasningen som gjengis øverst til høyre i Figur 4-2, viser at den teoretisk baserte modellen passer svært godt med rektorenes svar. Ladninger fra itemene mot de latente variablene, altså de alternative konstruktene, er også gode, der kun SC34Q19 lader under .6. En ladning på eksempelvis .8 mellom itemet og konstruktet indikerer at en endring på én standardavvik i itemet vil gi en endring på .8 standardavvik i den latente variabelen. Alle de fire konstruktene korrelerer betraktelig med hverandre, og er signifikante på .01-nivå, og den unike variansen til hvert av de fire konstruktene følgelig også tilsvarende lav. De betydelige korrelasjonene mellom konstruktene bekrefter samtidig at det var riktig å bruke oblimin rotasjon i de eksplorative faktoranalysene.

Figur 4-3 baserer seg på dataene fra de svenske rektorene. For å synliggjøre forskjellene mellom landene, er ladninger og korrelasjoner som avviker med mer enn .1 i positiv eller negativ retning i forhold til de norske verdiene markert med et henholdsvis grønt og rødt tall.

SVENSKE REKTORER



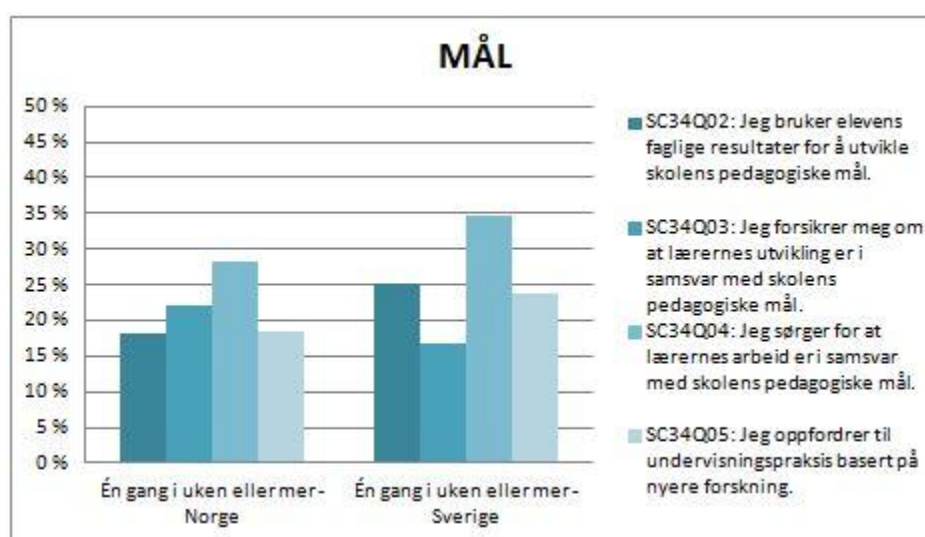
Figur 4-3: Konfirmerende faktoranalyse av alternative konstrukter med rektordata - Sverige

Figur 4-3 viser at modelltilpasningen også for Sveriges del er godt innenfor grenseverdiene til Hu og Bentler (1999). De to første konstruktene, og spesielt LÆRINGSMILJØ, har store likhetstrekk med den norske modellen, som indikerer at svarmønstrene til de norske og svenske rektorene samsvarer godt. Det samme kan ikke sies om de to siste konstruktene, og spesielt innenfor EVALUERING er det markante differanser. Korrelasjonene mellom konstruktene er også jevnt over er lavere, selv mellom de to første faktorene, som ellers er nokså like de norske.

4.3 Likheter og forskjeller mellom Norge og Sverige

→ Hvilke likheter og forskjeller mellom rektors mandat og handlingsrom i Norge og Sverige kan ledelseskonstruktene bidra til å belyse?

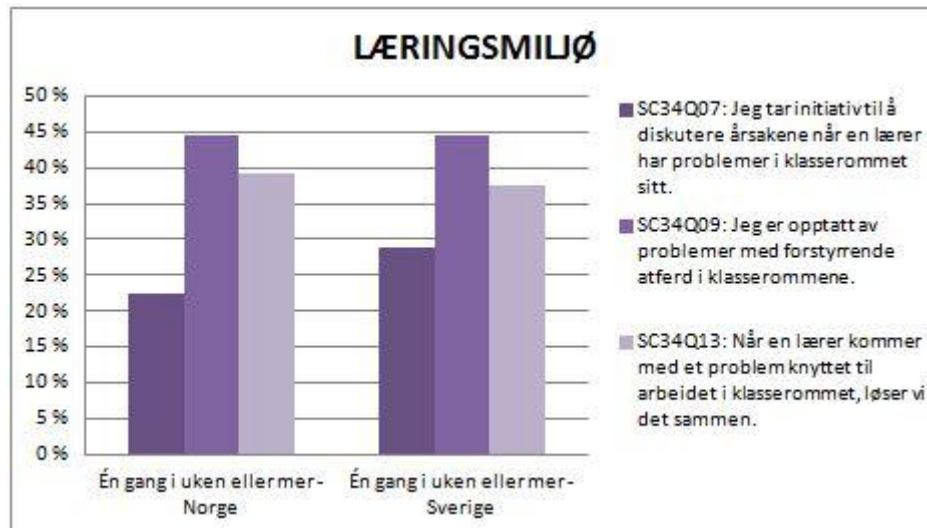
Figur 4-2 og 4-3 kan til en viss grad belyse relevante forskjeller mellom svarene fra de norske og svenske rektorene, som at svarene deres samsvarer mer i de to første konstruktene enn i de to siste. Derimot er det ikke nødvendigvis slik at ladningenes størrelse tilsvarende hvor høyt itemet er prioritert av rektorene, da de først og fremst angir itemets betydning for konstruktet. Gjennomgangen av likheter og forskjeller tilknyttet de ulike konstruktene vil innledes med en grafisk fremstilling av den prosentvise andelen rektorer som oppgir ”Én gang i uken” eller ”Mer enn en gang i uken” på de aktuelle itemene, da dette indikerer områder som i stor grad preger rektorenes arbeidshverdag. Det blir uoversiktlig å inkludere samtlige svarkategorier for samtlige item i en slik fremstilling, men i den påfølgende gjennomgangen vil relevante resultater fra hele frekvensanalysen trekkes inn. Frekvensanalysene baserer seg på rektordata, og en kontroll mot frekvensanalyser basert på flettede data gir avvik på under 1 prosentpoeng.



Figur 4-5: Rektorer som oppgir én gang i uken eller mer – MÅL

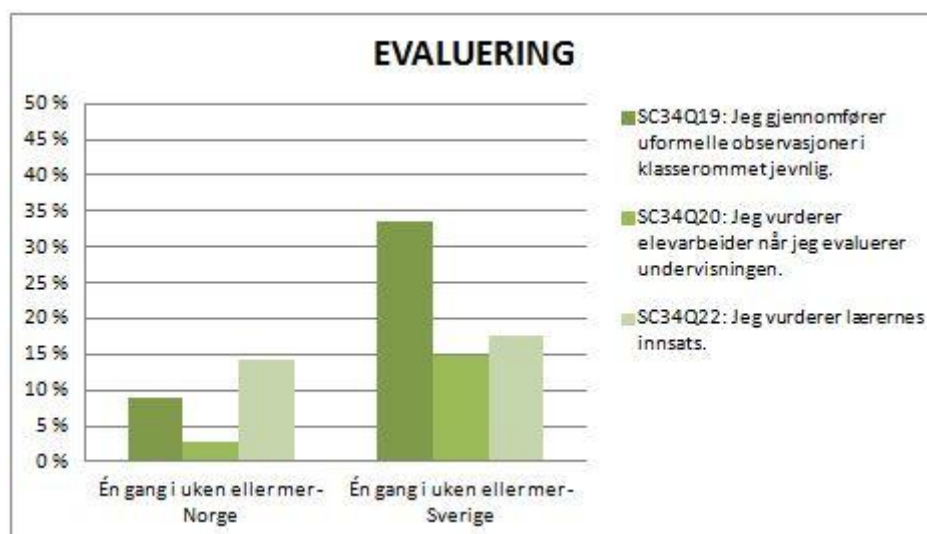
Som vi ser av Figur 4-5, stemmer bildet fra de konfirmerende faktoranalysene godt overens med det vi ser her. Det er noen mindre forskjeller mellom landene, og med unntak av SC34Q04 kan det virke som om svenske rektorer vektlegger målaspektet ved den pedagogiske ledelsen av skolen noe mer enn sine norske kolleger. Det å arbeide med skolens mål later til å være noe nær sagt alle rektorer bruker tid på i løpet av skoleåret, og for SC34Q02, SC34Q03 og SC34Q04 ligger andelen rektorer som oppgir at de aldri gjør dette på

1-2%, mens tallene for SC34Q05 er noe høyere, med 9% for norske rektorer og 3% for svenske rektorer. Reliabiliteten for konstruktet er høy for begge land, som indikerer at søylene innenfor hvert land inneholder mange av de samme rektorene, da itemene korrelerer sterkt.



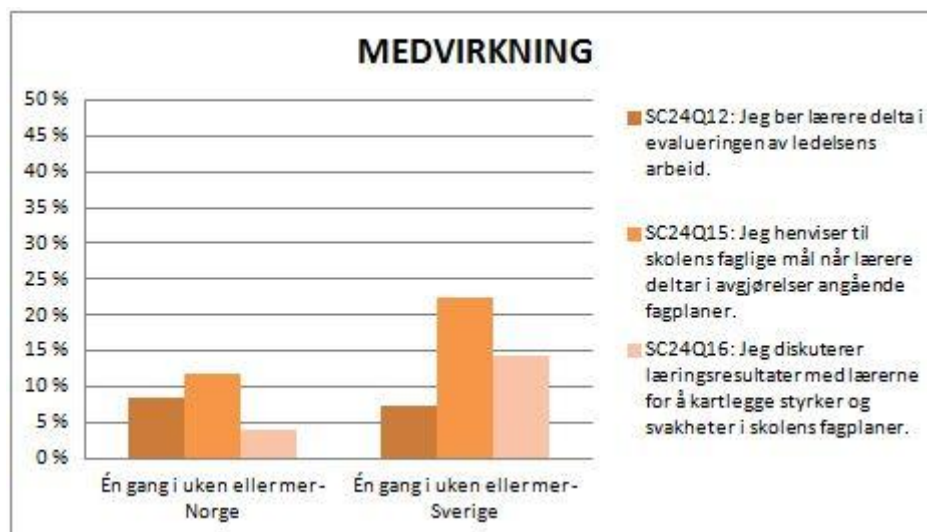
Figur 4-6: Rektorer som oppgir én gang i uken eller mer – LÆRINGSMILJØ

Figur 4-6 bekrefter at svenske og norske rektorer har svart nokså likt her, på tross av at norske rektorer ligger noe lavere når det kommer til SC34Q07, med 22% mot 29%. Det kan også nevnes at SC34Q09 og SC34Q13 topper rektorenes liste over ukentlige gjøremål blant samtlige item i de fire alternative konstruktene, og det kan virke som at arbeid tilknyttet LÆRINGSMILJØ i stor grad preger rektorenes hverdag som pedagogisk leder, eksemplifisert ved at hele 45% av rektorene oppgir at de minst ukentlig er ”opptatt med forstyrrede adferd i klasserommene”. Til slutt kan det nevnes at SC34Q07 og SC34Q13, til tross for en tilsvarende ordlyd, har en relativt ulik frekvens, spesielt for Norges del.



Figur 4-7: Rektorer som oppgir én gang i uken eller mer – EVALUERING

EVALUERING, delvis gjengitt i Figur 4-7 på forrige side, viser noen markante forskjeller mellom landene. Vi kan for eksempel se at nesten tre ganger så mange svenske som norske rektorer gjennomfører ukentlige observasjoner i klasserommene (SC34Q19). Noe som ikke kommer frem av figuren, er at 64% av norske rektorer oppgir at de gjør dette sjeldnere enn én gang i måneden, hvorav 20% ”Aldri”. De tilsvarende tallene for Sverige er 40% og 5%. At rektor vurderer elevarbeider (SC34Q20) er også mer utbredt i den svenske skolen, og hele fem ganger så mange svenske rektorer oppgir at de gjør dette ukentlig sammenlignet med Norge, der kun 3% har svart slik. Frekvenstabellene viser her i tillegg at 31% av norske rektorer oppgir at de aldri gjør dette, mot 18% i Sverige. Reliabiliteten for konstruktet havner for Norges del under .7, og det er grunn til å tro at SC34Q20 bidrar mest til dette, da rektorenes svar er sterkere forskjøvet mot den nedre delen av frekvenstabellen her enn for de to andre itemene. For Sveriges del er det derimot SC34Q19 som skiller seg mest fra de to andre, og som kan bidra til å forklare den lave reliabiliteten vi også finner her.



Figur 4-8: Rektorer som oppgir én gang i uken eller mer – MEDVIRKNING

MEDVIRKNING vektlegges også i noe større grad av svenske rektorer, selv om verdiene for SC34Q12 er så godt som like. Selv om Figur 4-8 ikke viser dette, er SC34Q12 også sterkere forskjøvet mot nedre ende av skalaen enn de to andre, som kan bidra til å forklare den forholdsvis lave reliabiliteten til konstruktet som helhet, selv om det for begge land er over .7. Et eksempel på dette, er at hele 16% av svenske rektorer aldri ”ber lærere delta i evalueringen av ledelsens arbeid”, mot 4% hos de norske.

4.4 Rektors ledelse og elevenes læring

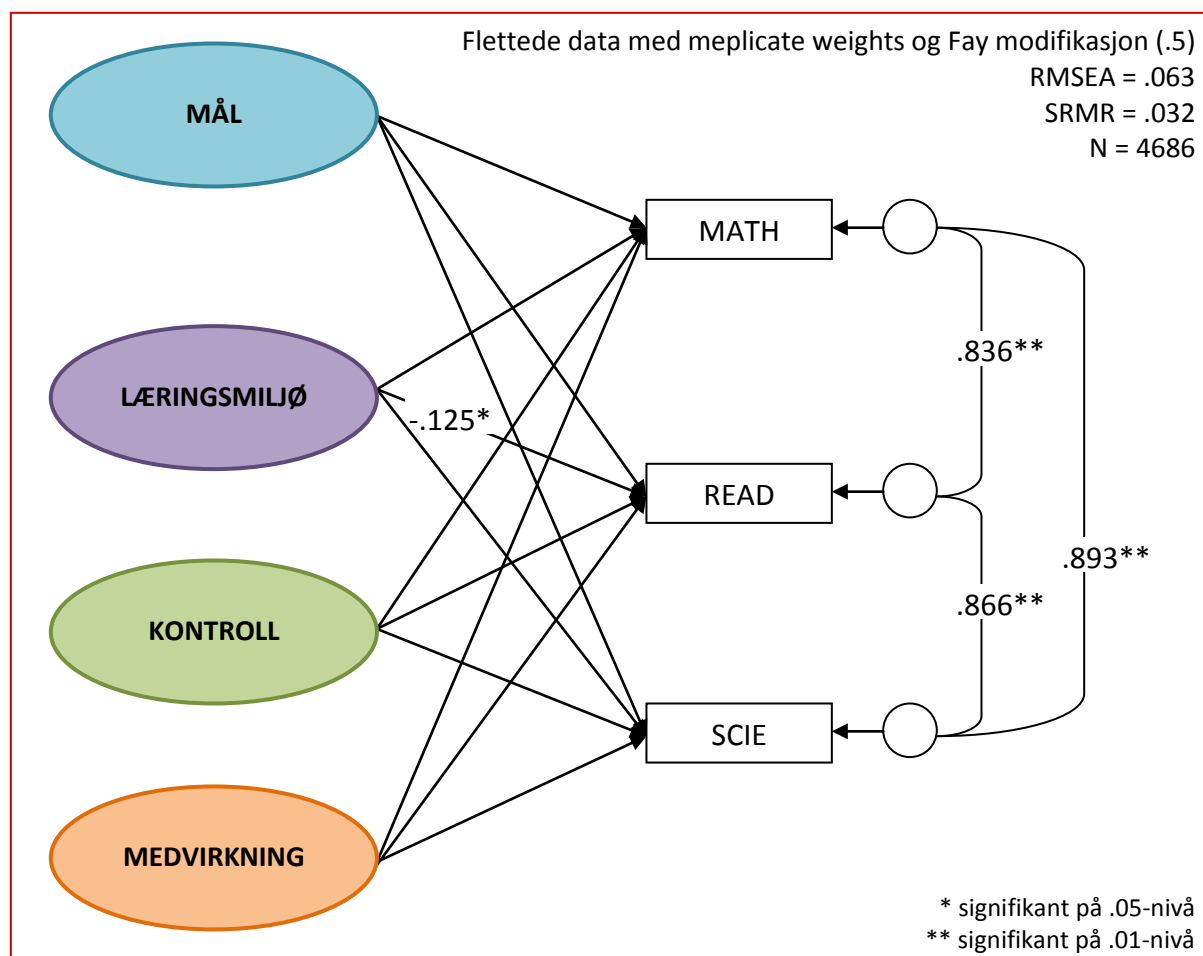
→ Kan man, ved å bruke strukturell modellering, finne en signifikant sammenheng mellom ledelseskonstruktene og elevenes resultater?

Den psykometriske analysen av PISA-konstruktene viste at det blir umulig å bygge en strukturell modell basert på disse, da modelltilpasningen var for svak. Det blir derfor de alternative konstruktene som må brukes for å belyse forskningsspørsmålet over.

4.4.1 Finnes det en signifikant sammenheng mellom de alternative konstruktene og elevenes læring?

Den strukturelle modellen nedenfor er en regresjonsmodell der man undersøker sammenhengen fra elevenes score innen de tre fagområdene på de alternative konstruktene. Itemene er av plasshensyn utelatt i modellene nedenfor, og kun signifikante verdier vises.

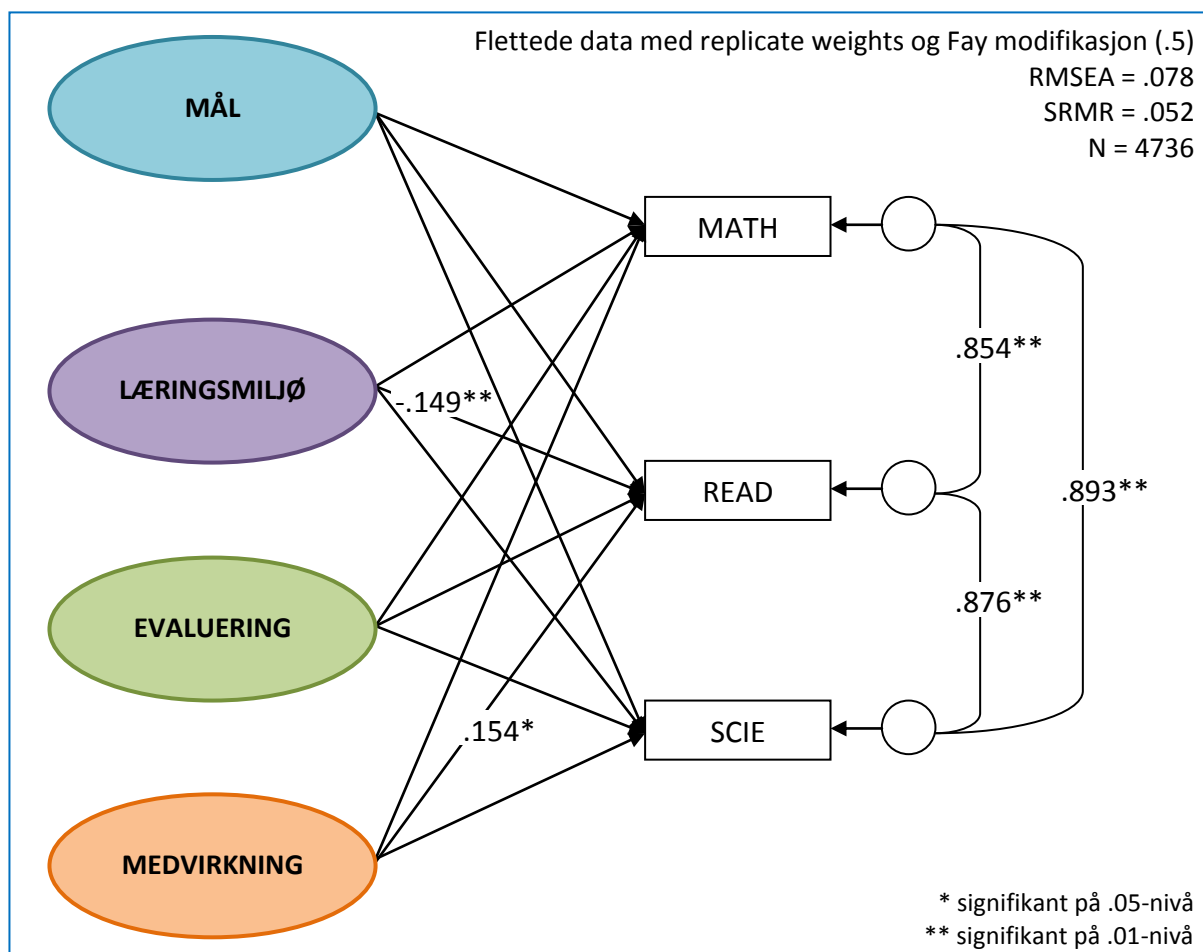
NORSKE REKTORER OG DERES ELEVER



Figur 4-9: Strukturell modell - Elevresultater regregert på de alternative konstruktene - Norge

Figur 4-9 viser oss at det for Norges del kun kan finnes én signifikant sammenheng, nemlig en svak negativ sammenheng fra READ på LÆRINGSMILJØ. Dette forteller oss at en positiv endring på én standardenhet i konstruktet LÆRINGSMILJØ medfører en negativ endring på .125 standardenheter i elevresultatene tilknyttet lesing. Videre indikerer den ene stjernen at denne sammenheng kun er signifikant på .05-nivå. Det er i denne sammenheng verdt å merke seg at Mplus7 ikke kan regne ut signifikansen for de standardiserte verdiene. Når sammenhengene her oppgis som signifikant på .05-nivå, er dette med andre ord basert på de ustandardiserte verdiene, men dette skal være uproblematisk i de aller fleste tilfeller. Som antydnet i forbindelse med den konfirmerende faktoranalysen, ser vi her at RMSEA har gått fra .026 til .063 mens SRMR faktisk blir noe bedre, og går fra .037 til .032. Denne studien vektlegger, på tross av dette usikkerhetsmomentet, at indikatorene fortsatt ligger godt innenfor grenseverdiene. Det siste aspektet ved modellen som skal nevnes før vi går videre, er at det som vi ser er en høy korrelasjon mellom de ulike fagområdene, noe som tilsier at elever som gjør det godt innen ett felt for det meste også gjør det godt innen de andre, og vice versa.

SVENSKE REKTORER OG DERES ELEVER

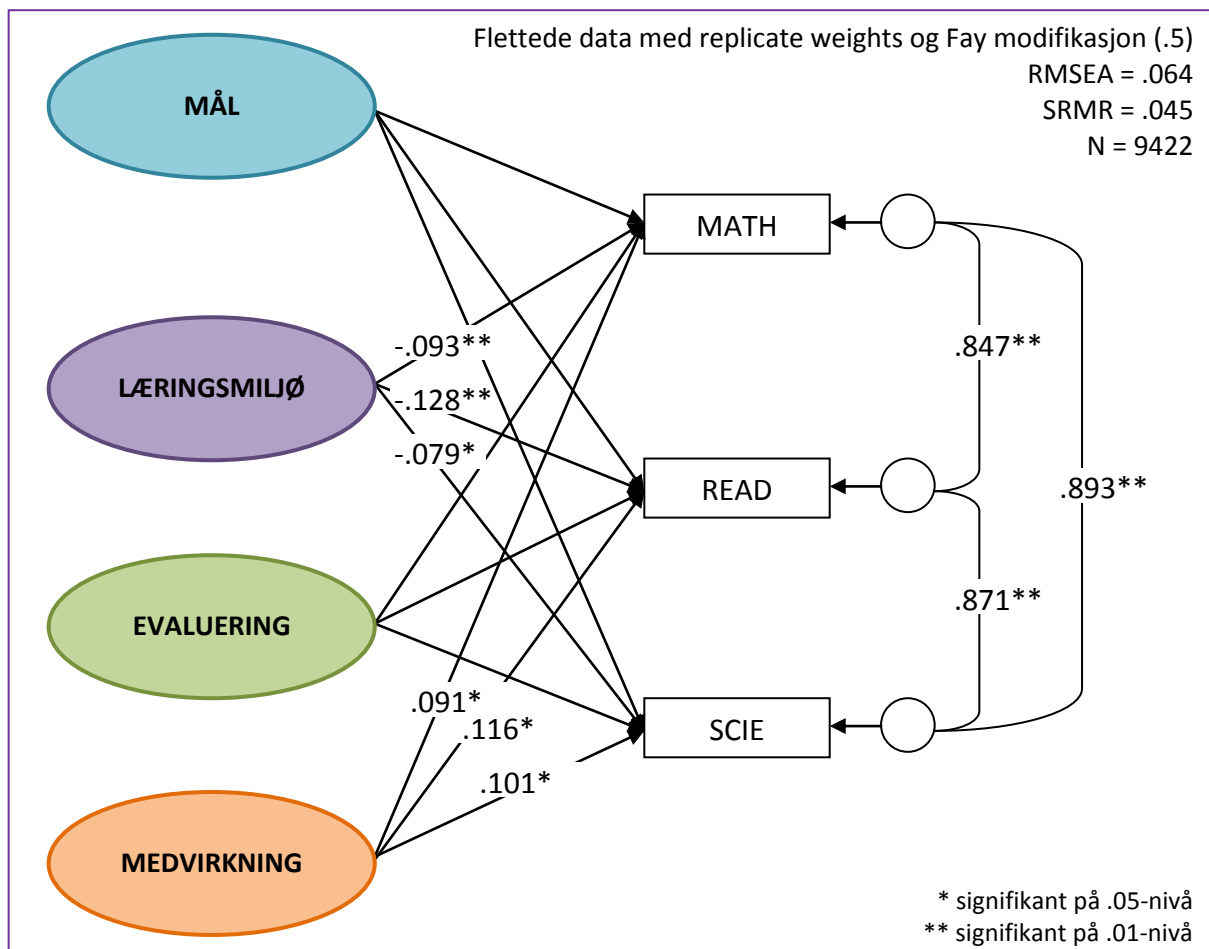


Figur 4-10: Strukturell modell - Elevresultater regregert på de alternative konstruktene - Sverige

Den svenske modellen i Figur 4-10 er i det store og det hele lik den norske, med noen unntak. Også her er det en svak sammenheng fra READ på MEDVIRKNING, dog noe sterkere og signifikant på 0.01-nivå. Videre er det også her sterke korrelasjoner mellom fagområdene. Fit-indikatorerne er noe svakere enn for den norske modellen, men både RMSEA og SRMR holder seg også i dette tilfellet innenfor grenseverdiene. Til forskjell fra den norske modellen, er også sammenheng fra SCIE på MEDVIRKNING signifikant, riktignok bare på .05-nivå.

Selv om N-verdiene for begge land ser høye ut, er det viktig å ha i bakhodet at antall rektorer fortsatt ligger på rundt to hundre for begge landene, da elevene ved samme skole har identiske data på variablene fra skole spørreskjemaet. Dette begrenser styrken på analysene, og for å øke den statistiske kraften kan de to datasettene derfor aggregeres i ett felles datasett, noe som er gjort nedenfor i Figur 4-11. RMSEA og SRMR er fortsatt innenfor grenseverdiene, og vi ser samtidig at doblingen av antall rektorer har ført til at vi har fått flere signifikante sammenhenger enn når landene ble analysert hver for seg.

AGGREGERT



Figur 4-11: Strukturell modell - Elevresultater regregert på de alternative konstruktene - Aggregert

Figur 4-11 viser de signifikante sammenhengene i det aggregerte datasettet med norske og svenske data, og resultatene herfra er i tråd med resultatene fra Figur 4-9 og 4-10. Det er verdt å merke seg at de svenske elevene har høyere vektinger enn de norske, da hver enkelt av de rundt 4700 svenske elevene som deltok representerer flere elever enn hver av de rundt 4700 norske. Dette innebærer at de svenske elevene bidrar mer i modellen, men betyr også at resultatene er representative for populasjonen av norske og svenske elever og deres rektorer.

Når det gjelder konstruktet LÆRINGSMILJØ, kan sammenhengen med READ gjenfinnes også her, og er med en verdi på $-.128$ signifikant på $.01$ -nivå. Det som derimot er nytt, er at de negative sammenhengene fra MATH og SCIE på LÆRINGSMILJØ, som ikke var signifikante i de to landene hver for seg, nå er blitt signifikante, og MATH på LÆRINGSMILJØ er også signifikant på $.01$ -nivå.

Noe tilsvarende kan ses tilknyttet MEDVIRKNING, der regresjonen fra READ på MEDVIRKNING har gått noe ned fra den svenske modellen, noe som ikke er overraskende sett i lys av fraværet av en tilsvarende signifikant sammenheng i den norske modellen. Videre har noen av de svakere sammenhengene også her nå blitt signifikante, men både MATH og SCIE på MEDVIRKNING er på $.05$ -nivå. Korrelasjonene mellom fagområdene er, ikke overraskende, høye også med aggregerte data, og er fortsatt signifikante på $.01$ -nivå.

På tross av større statistisk kraft i det aggregerte datasettet, ser vi heller ikke i Figur 4-11 noen signifikante sammenhenger tilknyttet de alternative konstruktene MÅL og EVALUERING.

5 DISKUSJON

I dette kapittelet skal trådene fra de foregående kapitlene samles gjennom at resultatene løftes frem og diskuteres i lys av teorien som er skissert i teorikapittelet. Inndelingen av dette kapittelet vil inneholde én del for hver av de fire forskningsspørsmålene, i tillegg til avsluttende delkapitler der studiens indre og ytre validitet, studiens bidrag til fagfeltet og tanker tilknyttet videre studier vil bli presentert.

5.1 Begrepsvaliditeten til PISA-konstruktene

→ Har konstruktene som inngår i PISA 2012 sin Index of Principal Management en tilfredsstillende begrepsvaliditet sett i lys av svarene til norske og svenske rektorer?

For å kunne besvare dette forskningsspørsmålet, må man, i tråd med Cook og Beckman (2006), undersøke om den *indre strukturen* til konstruktene er god, i tillegg til at de må ha et *innhold* som er solid fundamentert i relevant teori. Førstnevnte er allerede gjort, gjennom psykometriske analyser av reliabilitet og spesielt den konfirmerende faktoranalysen som ble presentert i resultatkapittelet. Sistnevnte vil gjøres gjennom en analyse av begrepsoperasjonaliseringen til itemene, der man vurderer konstruktens tilknytning til relevant teori. Da det til nå kun er gjennomført psykometriske analyser, vil disse gjennomgås raskt, før fokuset flyttes over på innholdsdimensjonen av begrepsvalideringen.

5.1.1 Hvordan er den indre strukturen til PISA-konstruktene?

Den overordnede konklusjonen tilknyttet PISA-konstruktens begrepsvaliditet er allerede presentert, da resultatdelen viste at konstruktene, på tross av å vise en akseptabel reliabilitet (se Tabell 4-2), totalt sett faller gjennom når de testes med konfirmerende faktoranalyser (se Tabell 4-3). Den teoretisk baserte modellen viste en uakseptabel modelltilpasning når den ble analysert med norske så vel som svenske data. Tilpasningen for de svenske dataene var riktignok ganske bra hvis man ser på RMSEA og SRMR alene, men CFI og TLI lå rimelig langt unna det som er anbefalt. Hvis man tok utgangspunkt i de flettede datafilene, der rektordataene analyseres i kraft av elevene sine, ble resultatet altså enda svakere. Dette vil forplante seg til enhver strukturell modell der man ønsker å kombinere rektordata og elevdata, og konstruktens uegnethet i forbindelse med strukturelle modeller reduserer det analytiske

potensialet til konstruktene dramatisk. Det er i denne sammenheng verdt å gjenta at kun ett av konstruktene, LEADCOM, hadde mer enn tre item, som er det som må til for at konstruktet er overidentifisert, med et positivt antall frihetsgrader. Dette gjorde det mulig å gjennomføre en frittstående konfirmerende faktoranalyse for dette konstruktet, men også her var modelltilpasningen for svak. Disse resultatene, i kombinasjon med at de andre konstruktene ikke kan etterprøves på samme måte, gjør at PISA-konstruktene heller ikke kan brukes i strukturelle modeller hver for seg. På tross av at konstruktene ikke kan brukes i forbindelse med elevens resultater, er det likevel både nødvendig og interessant å se litt nærmere på begrepsoperasjonaliseringen tilknyttet disse konstruktene, for å få et mer helhetlig bilde av begrepsvaliditeten. Denne delen av analysen kan også kaste lys over mulige områder der begrepsoperasjonaliseringen kan forbedres, selv om det ikke vil være mulig å anslå hvilken effekt svakhetene i begrepsoperasjonaliseringene kan ha hatt for de psykometriske analysene.

5.1.2 Hva kan sies om innholdet i PISA-konstruktene?

Instructional og transformational leadership har de siste tiårene ofte vært brukt for å beskrive motstykker innen skoleledelse, selv om Hallinger (2003) viser at lederstilene også har vesentlige fellestrekk. I denne delen vil de to lederstilene danne utgangspunkt for en vurdering av begrepsoperasjonaliseringen, som noen steder også vil gripe fatt i semantiske svakheter ved enkeltitem som kan være av betydning. Diskusjonen vil følgelig kaste lys over noen mulige områder der konstruktene og deres tilhørende item kan videreutvikles, sett i lys av at fremtidige PISA-undersøkelser sannsynligvis også vil forsøke å undersøke sider ved rektors pedagogiske ledelse gjennom skolespørreskjemaet.

LEADCOM

SC34Q02: Jeg bruker elevenes faglige resultater for å utvikle skolens pedagogiske mål.

SC34Q03: Jeg forsikrer meg om at lærernes utvikling er i samsvar med skolens pedagogiske mål.

SC34Q04: Jeg sørger for at lærernes arbeid er i samsvar med skolens pedagogiske mål.

SC34Q14: Jeg diskuterer skolens faglige mål med lærere på fellesmøter.

Framing and communicating the school's goals and curricular development (LEADCOM) inneholder altså fire item (se Tabell 4-2), der SC34Q02 og SC34Q04 kan finnes igjen i TALIS 2008, under fanen *Management school goal index* (se Vedlegg 7.2). SC34Q03 finner vi igjen i PISA 2009, men her finnes det som nevnt ingen underinndeling av itemene tilknyttet pedagogisk ledelse i rammeverket. SC34Q14 er ikke benyttet før i PISA eller TALIS, selv om det tematisk sett delvis er dekket av andre item i TALIS 2008. Alle fire itemene kan sies å

sortere under det OECD (2010b) kaller *Management of school goals*. (s. 35), som her regnes som en del av instructional leadership. Hvis man sammenligner med Hallingers verktøy PIMRS, er det i PIMRS hele 12 item tilknyttet "Frames goals" og "Communicates goals" (Hallinger, 1994, s. 7), noe som betyr at LEADCOM ikke vil kunne fange inn nyansene i dette aspektet ved rektors ledelse like godt som PIMRS. Før vi knytter konstruktet for tett opp til instructional leadership, er det viktig å huske at Hallinger (2003) også anser skolens mål som et viktig aspekt ved transformational leadership. Da itemene ikke sier noe om innholdet i skolens mål, er det vanskelig å tillegge det en av disse lederstilene alene.

SC34Q14, som altså er helt nytt i PISA 2012, skiller seg noe fra de tre første ved at man her drar inn lærerne som en aktiv samarbeidspartner i arbeidet med skolens mål, mens de tre første altså fokuserer på rektors arbeid. Slik sett kan det tenkes at dette itemet kan bidra til å forskyve konstruktet noe over fra instructional leadership mot transformational leadership, da en aktiv lærerstand er et mye tydeligere aspekt ved sistnevnte lederstil som helhet. Dette tematiske avviket kan også tenkes å svekke konstruktets reliabilitet som helhet, noe som følgelig ble undersøkt psykometrisk. Selv om dette strengt tatt faller inn under konstruktets indre struktur, og kunne vært håndtert i resultatdelen, tas disse analysene med her, da de springer ut av diskusjonen tilknyttet konstruktets innhold. Reliabilitetsanalysen av konstruktet uten SC34Q14 viser også at Cronbachs α bedres, som øker fra .796 til .858 når de norske dataene analyseres, og fra .780 til hele .900 med svenske data. Derimot vil en utelatelse av hvilket som helst av de andre itemene, ikke uventet, svekke reliabiliteten. LEADCOM ble som nevnt også testet med en CFA og funnet for lett, men det er ikke mulig å gjøre en individuell CFA av konstruktet uten SC34Q14, da konstruktet med kun tre uavhengige variabler ikke lenger vil være overidentifisert. Det er derimot mulig kjøre en CFA med alle PISA-konstruktene, men der SC34Q14 kuttet fra LEADCOM, noe som for Norges del faktisk fører til at RMSEA faller betydelig, fra .089 til .066, og SRMR fra .067 til .047. Dette betyr at indeksen som helhet nå faktisk havner innenfor Hu og Bentler (1999) anbefalinger for kombinasjonen av RMSEA og SRMR, der RMSEA ikke skal overstige .08, mens SRMR ikke skal overstige .06. Med flettede datafiler havner RMSEA derimot igjen godt utenfor grenseverdien, med en verdi på .099. De svenske dataene viser tilsvarende funn, og samlet sett ser det altså ut til at en utelatelse av SC34Q14, ikke bare for konstruktets del, men for indeksen som helhet, utgjør en markant forbedring, noe som bekreftes av en p-verdi på .00 i en tilhørende kjikvadratstest. Funnene her peker følgelig i retning av at det bør vurderes å kutte eller endre itemet i fremtiden, hvis konstruktet skal videreføres.

LEADINST

SC34Q05: Jeg oppfordrer til undervisningspraksis basert på nyere forskning.

SC34Q06: Jeg roser lærere som har elever som deltar aktivt i undervisningen.

SC34Q08: Jeg gjør lærere oppmerksomme på viktigheten av å utvikle elevenes kritiske og sosiale evner.

Instructional leadership (LEADINST) er et svært interessant konstrukt. Ingen av de tre itemene finnes i verken TALIS 2008 eller PISA 2009, og det er heller ikke lett å komme til bunns i hvorfor konstruktet har fått dette navnet. De tre itemene dreier seg riktignok om at rektor aktivt går inn og på ulike måter påvirker lærernes klasseromspraksis, men det kan være uheldig å bruke et allerede etablert begrep innen forskningsfeltet uten et tydelig teoretisk fundament, da det kan oppstå en begrepsforvirring rundt hva konstruktet er ment å måle. Hvis man skal analysere dette konstruktet i lys av Hallinger (2003) sin definisjon av *instructional leadership*, er det først og fremst tydelig at begrepet innbefatter langt flere aspekter ved rektors arbeid enn dette PISA-konstruktet kan sies å dekke. Dette understrekes av at *instructional leadership* i LEADINST undersøkes ved hjelp av 3 item, mot 66 for lederstilen som helhet i PIMRS (Hallinger, 1994, s. 7). Selv om ledelse av lærernes klasseromspraksis tydelig inngår i rektors funksjoner innenfor *instructional leadership*, ligger det i utgangspunktet ingen føringer angående denne funksjonens innhold. En formulering som ”viktigheten av å utvikle elevenes kritiske og sosiale evner” vil man nok kunne argumentere for at blir for spesifikt i denne sammenheng, og det er langt på vei uklart om dette itemet og konstruktet som helhet kan sies å ha en tydelig forbindelse til effektiv ledelse overhodet, selv om SC34Q05 altså har vist seg å passe godt inn i det alternative konstruktet MÅL. En misvisende bruk av et etablert begrep innen effektiv ledelse, som dette konstruktet nok er et eksempel på, kan altså i siste instans føre til at de publiserte PISA-publikasjonene feiltolkes og brukes i videre forskning på feilaktig grunnlag. Reliabiliteten for konstruktet ligger riktignok på .772 med norske data, mens den med svenske data ligger på .738, som også er akseptabelt, men det er ikke den indre strukturen som er dette konstruktets problem. Alt i alt kan dette konstruktet, hvis det skal videreføres, tjene på å at betegnelsen endres til noe som ikke er identisk med en veldefinert lederstil innen effektiv ledelse, samt at det presenteres et tydeligere teoretisk rammeverk som dette konstruktet skal innhente informasjon om.

LEADPD

SC34Q07: Jeg tar initiativ til å diskutere årsakene når en lærer har problemer i klasserommet sitt.

SC34Q09: Jeg er opptatt med forstyrrende adferd i klasserommet.

SC34Q13: Når en lærer kommer med et problem knyttet til arbeidet i klasserommet, løser vi det sammen.

Promoting instructional improvements and professional development (LEADPD) er basert på tre item som vi, på tross av noen små nyanser i ordleggingen, finner igjen i TALIS 2008. Ut ifra rammeverket finner vi også at de tre itemene tilhører det som i OECD (2010b) kalles *Instructional management index*, der rektor søker å forbedre lærernes undervisning. Det er likevel noen mulige svakheter i begrepsoperasjonaliseringen her, spesielt tilknyttet ordleggingen i SC34Q09: ”Jeg er opptatt med forstyrrende adferd i klasserommet.” Ordet *opptatt* kan bære i seg en tvetydighet tilknyttet om rektorenes frekvensangivelser skyldes en bevisst ledelsesstrategi eller om det er andre forhold ved skolen, herunder et utfordrende skolemiljø og disiplinære problemer, som fremtvinger at rektorene er opptatt med dette. Denne innsigelsen gjelder i og for seg også de to andre itemene, da en lett kan se for seg at disiplinære utfordringer ved en skole vil øke hyppigheten av disse ledelsesaktivitetene. Selv om PISA-undersøkelsene kan vise til at manglende disiplin i klasserommet har en negativ effekt på elevenes læring (OECD, 2013c), kan det for Norges del ikke påvises noen signifikante forskjeller i disiplinære problemer fra skole til skole (OECD, 2013b, s. 98). Sverige ligger derimot over OECD-gjennomsnittet på dette området, og det kan hende at dette, som vi skal se, kommer til uttrykk her. Det alternative konstruktet LÆRINGSMILJØ, som er identisk med LEADPD, viste seg nemlig å ha en negativ sammenheng med elevenes resultater innen samtlige fagområder i den aggregerte modellen (se Figur 4-11). Når man ser på landene hver for seg, ser vi videre at den negative sammenhengen med lesing fortsatt er signifikant, og at den er noe sterkere i den svenske modellen (se Figur 4-9 og 4-10). Disse funnene kan altså indikere at det er skolens disiplinære utfordringer heller enn rektors bevisste pedagogiske ledelse som måles i dette konstruktet, og kanskje spesielt i SC34Q09. Dette kan være et interessant innspill i forbindelse med utformingen av fremtidige skolespørreskjema.

Når det gjelder forbindelsen til instructional og transformational leadership, er fokuset på lærernes faglige utvikling en sentral del av rektors oppgaver. Selv om man som nevnt kan innvende at ikke alle problemer ”i klasserommet” eller ”knyttet til arbeidet i klasserommet” nødvendigvis har sin årsak i lærerens manglende pedagogiske eller didaktiske ferdigheter, kan man likevel argumentere for at både SC34Q07 og SC34Q13 er relevante i denne sammenheng. Forskjellen mellom itemene ligger primært i at det i SC34Q07 er rektor som tar

initiativet, mens det i SC34Q13 er læreren som gjør dette. Ettersom rektor er en langt mer aktiv deltaker innenfor instructional leadership, mens transformational leadership har et større fokus på lærerens personlige utviklingsmål, kan de to itemene muligens bidra til å skille mellom de to lederstilene, selv om en rektor som utøver instructional leadership neppe ville avslått et initiativ fra lærerens side, sett i lys av viktigheten av lærernes utvikling også her.

LEADTCH

SC34Q10: Jeg gir ansatte muligheten til å delta i beslutningsprosesser angående skolen.

SC34Q11: Jeg motiverer lærere til å utvikle en skolekultur basert på stadige forbedringer.

SC34Q12: Jeg ber lærere delta i evalueringen av ledelsens arbeid.

Teacher participation (LEADTCH) baserer seg som navnet tilsier på at lærere deltar i evaluerings- og beslutningsprosesser ved skolen, og som vi har sett i forbindelse med det tilsvarende alternative konstruktet, er det flere item som dreier seg om dette i spørreskjemaet. Dette konstruktet har, til forskjell fra de andre, en tydeligere tilknytning til transformational leadership enn instructional leadership, selv om det er viktig å påpeke at en rektor som ser på seg selv som transformational også vil ha et fokus på skolens mål som lærerens pedagogiske utvikling (Hallinger, 2003). Ingen av itemene er brukt i TALIS 2008 eller PISA 2009, og der SC34Q10 er eksplisitt på tilknytningen til lærermedvirkning, er medvirkningen i SC34Q12 mer indirekte, selv om det er naturlig at en rektor som inkluderer lærerne i evalueringen også vil ta hensyn til disse tilbakemeldingene i fremtidige beslutninger. SC34Q11 er enda vagere på dette punktet. Det er riktignok et viktig aspekt ved transformational leadership at rektor legger til rette for utvikling heller enn at rektor selv styrer denne utviklingen. Dette perspektivet tilsier muligens at itemet faller mer inn under transformational leadership enn under lærermedvirkning direkte, spesielt når man ser på hvilke andre item som ivaretar denne forbindelsen på en langt tydeligere måte.

Resultatdelen viste også at dette konstruktet, på tross av at reliabiliteten ligger rundt .70, faller helt fra hverandre i de eksplorerende faktoranalysene. At dette konstruktet kan ha godt av å bli revidert og eventuelt også revurdert fullstendig i fremtidige PISA-undersøkelser, understrekes av at vi ender opp med et alternativt konstrukt som gis navnet MEDVIRKNING, men der SC34Q12 er det eneste itemet som også inngår i LEADTCH. SC34Q11 viste seg derimot å ha en sterkere tilknytning til MÅL, men kun i den svenske faktorløsningen. SC34Q10 var ikke en del av faktorløsningen i noen av landene, og dette itemet vil bli berørt også senere, i forbindelse med begrepsvaliditeten til nettopp MEDVIRKNING.

5.2 Alternative ledelseskonstrukturer i PISA 2012

→ Kan man, basert på itemene tilknyttet pedagogisk ledelse i PISA 2012, etablere alternative ledelseskonstrukturer med en mer tilfredsstillende begrepsvaliditet?

Det ligger på mange måter et uutnyttet potensiale i PISA-undersøkelsene som basis for forskning på effektiv ledelse. Dette potensialet er både knyttet til at det tradisjonelt har vært liten interesse for denne formen for skolelederforskning i Norden, at skolespørreskjemaet med hell kan videreutvikles for å bedre begrepsvaliditeten, samt at det statistiske materialet allerede i dag har et uutnyttet statistisk potensiale. Det at PISA-konstruktene viste seg å falle noe gjennom, betyr nemlig ikke at de innsamlede dataene tilknyttet rektorenes pedagogiske ledelse er uten verdi, og det analytiske potensialet som ligger her er forsøkt utnyttet gjennom å etablere alternative ledelseskonstrukturer. De psykometriske analysene tilknyttet disse er allerede presentert, og innholdsdimensjonen tilknyttet konstruktene er delvis også berørt i resultatdelen, men de substansielle resonnementene vil utbroderes her.

5.2.1 Hvordan er begrepsvaliditeten til de alternative konstruktene?

MÅL

SC34Q02: Jeg bruker elevenes faglige resultater for å utvikle skolens pedagogiske mål.

SC34Q03: Jeg forsikrer meg om at lærernes utvikling er i samsvar med skolens pedagogiske mål.

SC34Q04: Jeg sørger for at lærernes arbeid er i samsvar med skolens pedagogiske mål.

SC34Q05: Jeg oppfordrer til undervisning basert på nyere forskning.

Som nevnt i resultatdelen er det tre av itemene i dette konstruktet som vi finner igjen i det tilsvarende PISA-konstruktet, og det er neppe hensiktsmessig å diskutere disse ytterligere her, da den tematiske tilknytningen til rektors arbeid med skolens mål er tydelig. SC34Q05 er på sin side allerede viet mye oppmerksomhet i resultatdelen, og resultatene med de tilhørende resonnementene der, dekker opp for de viktigste aspektene ved konstruktets begrepsvaliditet, også innholdsmessig. De psykometriske analysene viste at det var forsvarlig å inkludere itemet i konstruktet, og det ble videre antydning at det kan være rimelig å anta at en rektor som oppgir en høy frekvens på dette itemet, gjør dette nettopp fordi ”undervisning basert på nyere forskning” er et uttalt mål for skolens drift. Det er ikke mulig å etterprøve dette direkte, da rektorene ikke blir bedt om å spesifisere innholdet i skolens mål, men ved skoler som har et slikt mål, vil det å ”oppfordre til undervisning basert på nyere forskning” ha en tydelig

tilknytning til både SC34Q03 og SC34Q04, som begge fokuserer på lærernes arbeid og utvikling sett i lys av ”skolens pedagogiske mål”.

Når det gjelder tilknytningen til de to lederstilene, har de tre første itemene som nevnt en tilknytning til både instructional og transformational leadership, som begge vektlegger rektors arbeid med skolens mål. SC34Q05 kan derimot bidra til å forskyve itemet noe over mot instructional leadership, da rektor her inntar en aktiv rolle med tanke på å påvirke lærernes undervisningspraksis, noe som vil være unaturlig innenfor transformational leadership.

LÆRINGSMILJØ

SC34Q07: Jeg tar initiativ til å diskutere årsakene når en lærer har problemer i klasserommet sitt.

SC34Q09: Jeg er opptatt med forstyrrende adferd i klasserommet.

SC34Q13: Når en lærer kommer med et problem knyttet til arbeidet i klasserommet, løser vi det sammen.

Dette konstruktet er, som vi har sett, identisk med det PISA kaller *Promoting instructional improvements and professional development*. Her er det ikke like aktuelt med en diskusjon vedrørende om itemene hører sammen i et konstrukt eller ikke, da dette anses som nokså avklart. At konstruktet i resultatdelen ble tildelt et nytt navn, kan derimot drøftes ytterligere. Det nye navnet er tross alt ikke bare et resultat av at det kan være praktisk å skille PISA-konstruktene og de alternative konstruktene fra hverandre, men henger mer sammen med at PISA sin navngivning kan være noe problematisk. Dette er allerede berørt i diskusjonen tilknyttet PISA-konstruktet over, og det viktigste argumentet var at ”forstyrrende adferd i klasserommet”, som er utgangspunktet for SC34Q09, ikke nødvendigvis henger sammen med lærerens undervisning, slik PISA-konstruktets navn indikerer. Selv om dette i noen tilfeller sikkert kan stemme, stemmer det neppe at rektor primært har lærerens utvikling for øye hver eneste gang han er ”opptatt med” dette. Det at flere rektorer angir en høyere frekvens tilknyttet dette itemet enn de to andre i konstruktet, indikerer muligens også at rektor heller ikke alltid ser den forstyrrende adferden i sammenheng med læreren. Dette er hovedårsaken til at det alternative konstruktet er gitt navnet LÆRINGSMILJØ, da alle aktivitetene tilknyttet disse itemene har som mål å bedre læringsmiljøet i klasserommet, enten det er ved å fokusere på lærerens utvikling eller faktorer som ligger utenfor lærerens kontroll. Ulempen ved å navngi konstruktet slik, er at det ikke lenger har en like sterk tilknytning til et etablert begrep fra forskning på effektiv ledelse, men ett av argumentene ved navneendringen er jo også at denne tilknytningen kunne vært tydeligere.

I forbindelse med de to andre itemene i konstruktet, som på overflaten kan virke svært like, kan det være nyttig å trekke inn instructional og transformational leadership. For mens både SC34Q07 og SC34Q13 dreier seg om en lærer som har problemer i klasserommet, ligger det en viktig nyanse tilknyttet hvem som er initiativtaker i situasjonen, henholdsvis rektor og læreren. Sånn sett kan man spore en sterkere grad av instructional leadership i førstnevnte, da denne bærer i seg et ovenfra-og-ned-perspektiv som kjennetegner nettopp denne lederstilen (Hallinger, 2003). SC34Q13 er derimot utformet på en måte som tilsier at enhver rektor som er opptatt av et godt læringsmiljø, enten dette involverer lærernes profesjonelle utvikling eller ikke, vil kunne stille seg bak dette. Man kan også ane en mulig svakhet ved konstruktet tilknyttet at det, som det eneste av de alternative konstruktene, har en svak negativ sammenheng med elevenes læring, som allerede er diskutert i tilknytning til LEADPD over. Det er i utgangspunktet vanskelig å se for seg at en rektors fokus på lærernes pedagogiske utvikling skal kunne ha en sammenheng med svakere resultater. Derimot er det langt lettere å være med på et resonnement som tilsier at en rektor med store disiplinære problemer ved skolen kan oppleve at elevene presterer svakere. Hvis man gjennomfører en regresjon med elevenes resultater på enkeltitemene i konstruktet hver for seg, finner man også at det kun er SC34Q09 som har en signifikant negativ sammenheng med de tre fagområdene, noe som altså kan tyde på at dette itemet, og konstruktet som helhet, fanger opp disiplinære problemer i vel så stor grad som bevisst pedagogisk utvikling.

EVALUERING

SC34Q19: Jeg gjennomfører uformelle observasjoner i klasserommet jevnlig,

SC34Q20: Jeg vurderer elevarbeider når jeg evaluerer undervisningen.

SC34Q22: Jeg vurderer lærernes innsats.

I Tabell 4-4 og 4-5 kan man se at dette konstruktet var en del av faktorløsningen både med svenske og norske data, men det er altså ikke et konstrukt som benyttes i noen av PISA-rapportene. Dette kan ha sammenheng med at det, som vi kan se av Vedlegg 7.1, av ukjente årsaker på et tidspunkt har blitt fjernet et item i denne delen av spørreskjemaet (SC34Q21), noe som kan ha medført at man valgte å ikke bruke konstruktet i PISA-rapportene. Itemene har nemlig noen tydelige fellestrekk, ved at de alle dreier seg om at rektor skaffer seg en oversikt over og evaluerer skolens virksomhet ved å observere i klasserommet, som kan involvere både elever og lærere, ved å vurdere elevarbeider og ved å vurdere lærerne. Denne aktive evalueringen henger tydeligst sammen med instructional management, jfr. Tabell 2-1, og kan ha hatt sitt utgangspunkt i item fra PIMRS-konstruktet *Supervision/Evaluation* og/eller

Monitors student progress, som består av henholdsvis 11 og 7 item (Hallinger, 1994, s. 7). Vi finner riktignok ikke igjen noen av disse itemene i TALIS 2008, så forbindelsen til PIMRS er således uklart. Det kan også være verdt å merke seg at reliabiliteten til dette konstruktet er nokså svak for begge landene, og faller for Norges del faktisk under .70. Hvis denne svake reliabiliteten gjenspeiler seg i mange land, kan også dette være en grunn til at konstruktet ikke er rapportert i PISA-publikasjonene. En siste innvending, forholder seg til at ordet ”jevnlighet”, som i seg selv sier noe om hyppighet, er en del av formuleringen i SC34Q19. Dette kan kollidere med at rektorene skal angi en frekvens, selv om det er vanskelig å se hvordan dette skulle føre til at rektor svarer annerledes enn hvis ordet var utelatt. Det er som vi ser grunn til å være skeptisk til dette konstruktet, men når det likevel er tatt med i de strukturelle modellene, skyldes dette følgende: Den totale modelltilpasningen er fortsatt svært god, forskjellen mellom modellene er ikke signifikant med tanke på endringen i antall frihetsgrader, og konstruktet bærer i seg et interessant perspektiv tilknyttet rektors pedagogiske ledelse.

MEDVIRKNING

SC34Q12: Jeg ber lærere delta i evalueringen av ledelsens arbeid.

SC34Q15: Jeg henviser til skolens faglige mål når lærere deltar i avgjørelser angående fagplaner.

SC34Q16: Jeg diskuterer læringsres. med lærerne for å kartlegge styrker og svakheter i skolens fagplaner.

Som vi så i resultatdelen, var det ikke bare enkelt å forene de to faktorløsningene i et felles konstrukt her, men slik det står nå, har konstruktet en akseptabel indre struktur, og alle itemene fokuserer på at lærere involveres og medvirker i prosesser tilknyttet den pedagogiske ledelsen av skolen, enten dette er gjennom deltakelse i beslutningsprosessene eller gjennom evalueringen av skolens virksomhet. Konstruktet bærer også i seg et viktig aspekt ved transformational leadership, som preges av at ansvaret for utviklingsprosessene i større grad kan spres i kollegiet, at det i større grad legges til rette for endring nedenfra-og-opp og at lærernes mål langt på vei sidestilles med skolens overordnede mål. Tilknytningen til instructional leadership er likevel til stede, da rektors involvering i arbeid tilknyttet skolens ”fagplaner”, som nevnes i både SC34Q15 og SC34Q16, er et sterkere fokusområde innenfor instructional leadership. Selv om disse tre itemene altså dekker ulike aspekter ved lærernes medvirkning i skolens interne prosesser, er det verdt å legge merke til at det finnes item som kunne og kanskje burde sortert under dette konstruktet, men som ikke gjør det. SC34Q14 er allerede nevnt, men spesielt SC34Q10; ”Jeg gir ansatte muligheten til å delta i beslutningsprosesser angående skole.”, bærer jo langt på vei i seg selve essensen av

konstruktet, men frekvensene som rektorene oppgir for dette itemet er langt høyere enn for de tre som inngår i itemet, noe som forklarer den manglende tilhørigheten til konstruktet. Forklaringen på dette henger nok sammen med at dette itemet beskriver medvirkning på et mer generelt nivå, og man kan nærmest si at de tre itemene i konstruktet sorterer under dette itemet like mye som de sorterer under konstruktet.

5.3 Likheter og forskjeller mellom Norge og Sverige

→ Hvilke likheter og forskjeller mellom rektors mandat og handlingsrom i Norge og Sverige kan ledelseskonstruktene bidra til å belyse?

Nå som de alternative konstruktene er gjennomgått, er det på sin plass å se litt nærmere på hvordan rektorenes svar kan bidra til å belyse likheter og ulikheter mellom rektors mandat og handlingsrom i de to landene. Konstruktene vil først diskuteres hver for seg, og vil trekke veksel på resultater fra reliabilitetsanalysene, de konfirmerende faktoranalysene, frekvensfremstillingene og de strukturelle modellene, i tillegg til relevant teori. Deretter vil de norske og svenske dataene diskuteres samlet, der diskusjonen tilknyttet konstruktene ses i lys av det bildet de konfirmerende faktoranalysene gir oss av konstruktene samlet sett.

5.3.1 MÅL

SC34Q02: Jeg bruker elevenes faglige resultater for å utvikle skolens pedagogiske mål.

SC34Q03: Jeg forsikrer meg om at lærernes utvikling er i samsvar med skolens pedagogiske mål.

SC34Q04: Jeg sørger for at lærernes arbeid er i samsvar med skolens pedagogiske mål.

SC34Q05: Jeg oppfordrer til undervisning basert på nyere forskning.

Dette konstruktet preges først og fremst av at samtlige item har en nokså sentral plass i mange norske og svenske rektors bevissthet, for selv om andelen rektorer som oppgir at de bruker tid på disse oppgavene hver uke er beskjeden, er det så godt som ingen som ikke setter av tid til å jobbe med og følge opp skolens pedagogiske mål i løpet av skoleåret. Målstyring er altså ikke bare blitt et instrument staten styrer skolene med, men også et instrument rektor bruker i sin pedagogiske ledelse av skolen. Det kan igjen være greit å merke seg at innholdet i disse målene ikke fanges opp i skole spørreskjemaet, og at det således er vanskelig å få ytterligere kunnskap om rektors prioriteringer tilknyttet skolens mål. I resultatdelen nevnes det også at tre av fire item har en noe høyere tilslutning blant svenske rektorer, og selv om forskjellene ikke er store, kan vi like fullt spore en viss forskjell. Det er også interessant at det itemet som

tydeligst fokuserer på lærernes utvikling (SC34Q03) er det eneste der norske rektorer ligger høyere, mens itemene der rektor utvikler skolens mål (SC34Q02), sørger for at skolens mål følges opp av lærerne (SC34Q04) og aktivt oppfordrer til konkrete utviklingstiltak (SC34Q05) alle ligger høyere for Sveriges del. Selv om dette på langt nær er tilstrekkelig for å konkludere med at svenske rektorer er sterkere preget av tankegods fra instructional leadership, er det likefullt et innspill som kan tas med i den videre diskusjonen.

5.3.2 LÆRINGSMILJØ

SC34Q07: Jeg tar initiativ til å diskutere årsakene når en lærer har problemer i klasserommet sitt.

SC34Q09: Jeg er opptatt med forstyrrende adferd i klasserommet.

SC34Q13: Når en lærer kommer med et problem knyttet til arbeidet i klasserommet, løser vi det sammen.

Det er allerede påpekt at norske og svenske rektorer er svært samstemte i sine frekvensangivelser tilknyttet dette konstruktet, selv om reliabiliteten er noe høyere med svenske data. SC34Q09 og SC34Q13 er også de to itemene som har de høyeste frekvensangivelsene av itemene som inngår i de fire konstruktene, noe som viser at rektorer i begge land bruker mye tid på å følge opp problemer i klasserommet, og at mange rektorer opplever dette som en sentral del av sitt virke. På item SC34Q07, der rektor inntar en mer aktiv rolle, ligger svenske rektorer igjen noe høyere enn de norske, der 29% ”tar initiativ til å diskutere årsakene når en lærer har problemer i klasserommet sitt” ukentlig, mens dette en altså gjelder kun 22% av de norske rektorene. Forskjellen tilsvarer den vi fant tilknyttet itemene fra MÅL der rektor inntar en noe mer aktiv rolle, noe som kan tyde på at trekk fra instructional leadership også i dette tilfellet har fått et noe sterkere fotfeste i Sverige. Et bidrag til den noe høyere tilslutningen til SC34Q07 kan muligens også tilskrives at svensk skole er noe mer plaget av uro i undervisningen enn den norske (Kjærnsli og Jensen, 2013, s. 263), men dette gjenspeiles ikke i de andre to itemene, og Norge ligger faktisk marginalt høyere på SC34Q13, som representerer den lærerinitierte varianten av SC34Q07.

5.3.3 EVALUERING

SC34Q19: Jeg gjennomfører uformelle observasjoner i klasserommet jevnlig,

SC34Q20: Jeg vurderer elevarbeider når jeg evaluerer undervisningen.

SC34Q22: Jeg vurderer lærernes innsats.

Til forskjell fra de to foregående konstruktene, får vi i dette konstruktet se betydelige forskjeller mellom norske og svenske rektorer. Dette er spesielt tydelig i tilknytning til at

svenske rektorer i langt større grad "...gjennomfører uformelle observasjoner i klasserommet..." enn det norske rektorer gjør. Forskjellen som vises i Figur 4-7 tilknyttet SC34Q19 forplanter seg som påpekt i resultatdelen videre nedover i alle svaralternativene, helt ned til kategorien "Aldri", der 20% av norske rektorer oppgir at de hører hjemme, mot bare 5% av svenske rektorer. Om vi i de to foregående konstruktene så en tendens til en noe høyere tilstedeværelse av instructional leadership i svenske skoler, så forsterker dette konstruktet definitivt dette inntrykket. Fokuset på evalueringsaspektet ved rektors ledelse kan ha sin bakgrunn i et sterk fokus på ansvarliggjøring i den svenske skolen på 2000-tallet, et aspekt ved skolen vi har sett kom tidligere og har utviklet seg lenger i Sverige enn i Norge. Hallerström (2006) konkluderer også med at det "å være ute i virksomheten" på 2000-tallet har blitt så nært assosiert med pedagogisk ledelse at det nærmest har blitt en norm for rektorene, og sammenligner denne formen for lederskap med nettopp instructional leadership (s. 156). EVALUERING har som nevnt også klare likhetstrekk til konstrukt vi finner igjen i PIMRS, som jo nettopp er konstruert for å kartlegge denne lederstilen. Figur 4-7 viser også at svenske rektorer i vesentlig større grad enn norske "vurderer elevarbeider når jeg evaluerer undervisningen" (SC34Q20), og vi har sett at nesten en tredjedel (31%) av norske rektorer oppgir at de aldri gjør dette, mot 18% i Sverige, som riktignok også er en betydelig andel. Selv om dette kan være et tegn på at dette er et område som prioriteres nokså lavt, er det på sin plass å påpeke at itemene tilknyttet pedagogisk ledelse kun kartlegger hva rektor bruker tiden sin på, og det kan således tenkes at rektor har delegert bort oppfølgingen av oppfølgingen av elevarbeider til andre personer i skolens ledelse. Uavhengig av dette, er det like fullt ikke urimelig å ta dette som ytterligere et eksempel på en større grad av instructional leadership blant svenske rektorer. Når det kommer til "å vurdere lærernes innsats" (SC34Q22), ligger de norske og svenske rektorene derimot nokså likt. Det er likevel verdt å notere at det som vurderes, nemlig "lærernes innsats", ikke har en like tydelig tilknytning til undervisningen, og følgelig instructional leadership, som de to andre itemene hadde, på tross av at det også her dreier seg om vurdering av skolens virksomhet. Avslutningsvis er det riktignok på sin plass å påpeke at man skal være litt forsiktig med å legge for stor vekt på konklusjoner som tar utgangspunkt i konstruktet som en helhet, spesielt for Norges del, da det har en forholdsvis lav reliabilitet, mens den med svenske data ligger over .7. På den andre siden kan den lave reliabiliteten for Norges del tolkes som nok et signal om at instructional leadership i liten grad er en del av norske rektors bevissthet.

5.3.4 MEDVIRKNING

SC34Q12: Jeg ber lærere delta i evalueringen av ledelsens arbeid.

SC34Q15: Jeg henviser til skolens faglige mål når lærere deltar i avgjørelser angående fagplaner.

SC34Q16: Jeg diskuterer læringsres. med lærerne for å kartlegge styrker og svakheter i skolens fagplaner.

Gjennomgangen av resultatene tilknyttet Figur 4-8 viste at dette konstruktet som helhet også vektlegges i noe større grad av svenske rektorer, som er interessant, da dette er det konstruktet som i størst grad kan sies å sortere under transformational leadership. Dette perspektivet dempes, som nevnt, noe av fokuset på skolens fagplaner i SC34Q15 og SC34Q16, som også er de itemene som de svenske rektorene prioriterer høyere enn de norske. Det kan riktignok tenkes at det noe større fokuset på læreplanarbeid og skolens fagplaner ikke nødvendigvis er en bevisst ledelsesstrategi fra de svenske rektorenes side, men at det rett og slett kan ha sin bakgrunn i at den svenske grunnskolen fikk nye læreplaner i 2011. Mens det i 2012 altså var seks år siden Kunnskapsløftet ble innført i Norge, var den svenske læreplanen bare ett år gammel. Prosessene tilknyttet nye læreplaner er omfattende og tidkrevende (Aasen et al., 2012, s. 295), og det er således ikke uventet at svenske rektorer satte av mer tid til arbeid med skolen fagplaner enn de norske. Selv om SC34Q10 ikke er en del av konstruktet, kan det i denne sammenheng likevel være på sin plass å se raskt på hva svenske og norske rektorer svarer på spørsmålet om de ”gir ansatte muligheten til å delta i beslutningsprosesser angående skolen”. Dette itemet bærer langt på vei i seg essensen av medvirkningsperspektivet tilknyttet rektors pedagogiske ledelse, og her finner man at 50% av svenske rektorer oppgir én gang i uken eller mer her, mot 46% i Norge. Når det gjelder SC34Q12, er det derimot verdt å merke seg at 16% av svenske rektorer oppgir at de aldri ”ber lærerne delta i evalueringen av ledelsens arbeid, mot 4% i Norge. Den noe høyere tilstedeværelsen av rektorer som ikke lar lærerkollegiet evaluere ledelsen, kan tas til inntekt for en høyere tilstedeværelse av instructional leadership i den svenske skolen, noe som jo stemmer overens med bildet vi har sett så langt. I det store og det hele er det likevel viktig å påpeke at det er lite som tyder på at medvirkningsperspektivet står svakere i den svenske skolen enn i den norske ut over dette.

5.3.5 ET SAMLET PERSPEKTIV

Som gjennomgangen av norsk og svensk skoles utvikling siden 70-tallet viste, er det fortsatt store likheter tilknyttet rektors mandat og handlingsrom generelt, og som pedagogisk leder. Dette dreier seg spesielt om tilstedeværelsen av ”den skjulte kontrakten” i begge land, som fortsatt i betydelig grad begrenser rektors handlingsrom (Johansson og Bredeson, 2011;

Møller, 2011). På tross av åpenbare likheter, viste den historiske gjennomgangen at det er andre sider ved rektors mandat og handlingsrom som i dag fremstår som mer fremtredende i svensk skole. For det første kunne det vises til flere OECD-rapporter der det ble konkludert med at svenske rektorer hadde, og fortsatt har, en noe høyere grad av autonomitet i sin stilling (OECD, 2008a, 2014a). Dette hører delvis sammen med at de er under et sterkere press med tanke på ansvaret for skolens resultater, noe som forsterkes ytterligere i delene av landet der konkurrerende private skoler og markedsprinsipper er mest fremtredende (Lundahl et al., 2013). Noe som også kan spille inn i denne sammenheng, er at rektorutdanningen ikke bare er mer utbredt, men at den også har et sterkere fokus på rektorenes ulike ledelsesfunksjoner enn den norske rektorutdanningen, som er mer verdiorientert (Hybertsen Lysø et al., 2011).

Spørsmålet blir så om dette også kommer til syne i rektorenes svar. Som nevnt finnes det flere argumenter for at spesielt konstruktet EVALUERING viser at svenske rektorer har et tydeligere fokus på selv å observere og evaluere undervisningen. I tillegg kan vi i Sverige se antydninger til et noe mer rektorfokusert arbeid med skolens MÅL, et mer aktivt standpunkt i forhold til lærere med problemer i klasserommet i konstruktet LÆRINGSMILJØ og en noe høyere andel av rektorer som styrer skolen uten at lærerkollegiet får evaluere ledelsens arbeid i MEDVIRKNING. Vi kan altså i samtlige konstruktter finne tegn som peker mot at visse aspekter ved instructional leadership i noe større grad har fått fotfeste i svenske skoler sammenlignet med norske, selv om det neppe finnes grunnlag her for å si at dette preger den svenske skolen som helhet. Sammenlignet med norske rektorer, virker det heller ikke som at instructional leadership har erstattet tilstedeværelsen av lærermedvirkning, som er fremtredende i begge land.

Når det gjelder rektors autonomi i Sverige, kan det faktum at korrelasjonene mellom de alternative konstruktene er lavere i Figur 4-3 sammenlignet med Figur 4-2 tyde på at svenske rektorer er noe mer rendyrket i sine prioriteringer tilknyttet pedagogisk ledelse, og at de i noe mindre grad enn de norske rektorene gjør litt av det ene og litt av det andre. Dette kan tolkes som et tegn på rektor står noe mer fritt i sin rolle på skolen, som kan være et resultat av at svenske rektorer generelt er noe mer autonome enn sine norske kolleger. Det totale bildet av rektors pedagogiske ledelse fremstår følgelig, på tross av at korrelasjonene også for Sverige er betydelige, som noe mindre komplekst enn det norske, som i større grad preges av høye korrelasjoner mellom konstruktene. I det store og det hele er likhetene mellom de to landene

allikevel langt mer påfallende enn forskjellene dem imellom, noe som gjenspeiler det bildet vi fikk av utdanningssystemet som helhet.

5.4 Pedagogisk ledelse og elevenes resultater

→ Kan man, ved å bruke strukturell modellering, finne en signifikant sammenheng mellom ledelseskonstruktene og elevenes resultater?

Et viktig mål med denne studien har vært å undersøke det analytiske potensialet tilknyttet rektorenes ledelse som ligger i PISA 2012 ved å undersøke begrepsvaliditeten til konstruktene tilknyttet PISA 2012 sin Index of School Management og undersøke om det er mulig å finne andre konstrukt med god indre struktur og et solid teoretisk fundament. Et annet formål med studien tar derimot ikke kun utgangspunkt i å undersøke det analytiske potensialet, men også å utnytte dette, mer spesifikt gjennom å se om det finnes en signifikant sammenheng mellom rektors pedagogiske ledelse og elevenes resultater. Før disse funnene diskuteres ytterligere, kan det riktignok være på sin plass å kort adressere kritikken som tar utgangspunkt i at PISA-undersøkelsene er for snevre i hva de måler, og at undersøkelsene tillegges for stor vekt i utdanningspolitikken. Denne debatten er det ikke mulig å presentere fullt ut her, og denne studien har også som utgangspunkt at det ikke er essensielt for funnenes gyldighet om PISA gir et utfyllende bilde av de mål som skolen skal jobbe mot. Det er videre vanskelig å se at elevenes resultater kan avfeies som ikke relevante på generell basis, da kompetansene som testes i PISA er såpass brede at de dekker mange sider av de norske og svenske læreplanene.

5.4.1 Hva forteller de signifikante sammenhengene mellom de alternative konstruktene og elevenes resultater oss?

Som Figur 4-9, 4-10 og 4-11 viser, er de signifikante sammenhengene mellom de alternative konstruktene og elevenes resultater ikke bare relativt få, men også så svake at det er svært vanskelig å skulle dra noen slutninger på bakgrunn av disse. Den eneste sammenheng som vi finner i begge landene hver for seg, som også er den eneste som er signifikant for Norges del, er en svak negativ sammenheng mellom LÆRINGSMILJØ og READ. Dette indikerer at skoler som angir en høyere frekvens tilknyttet dette konstruktet, generelt sett har noe svakere resultater innenfor lesing. Hvis man analyserer den aggregerte datafilen, som vises i Figur 4-11, ser vi at konstruktet LÆRINGSMILJØ faktisk utviser en negativ sammenheng med alle

de tre fagområdene, uten at det finnes noen grunnlag for å konkludere om hvorfor det er slik. Hvis man går videre inn og utfører en regresjonsanalyse av elevresultatene på de tre itemene som utgjør konstruktet hver for seg, finner man at bidraget fra SC34Q09 er det eneste som er signifikant i seg selv. Den potensielle problematikken tilknyttet hva dette itemet og konstruktet som helhet faktisk måler, er allerede nevnt, og tar altså utgangspunkt i at de svakere elevresultatene ikke nødvendigvis har en sammenheng med rektors pedagogiske ledelse, men heller er et resultat av generelle disiplinære problemer ved skolen. Slik sett er det mye som tyder på at denne sammenhengen, som uansett er svak, ikke er videre relevant i en diskusjon om betydningen av rektors pedagogiske ledelse for elevenes resultater.

De tre signifikante funnene som viste en positiv sammenheng mellom rektors pedagogiske ledelse og elevenes resultater, var også svært svake, og er alle relatert til konstruktet MEDVIRKNING. Ingen av disse sammenhengene var signifikante når kun de norske dataene ble analysert, mens sammenhengen fra READ på MEDVIRKNING er signifikant med svenske data. Også her er det mulig å utføre en regresjon av elevresultatene på de tre itemene i konstruktet hver for seg, men det aggregerte datasettet viser kun én signifikant sammenheng, fra SCIE på SC34Q12. Selv om disse funnene muligens kan danne et utgangspunkt for videre undersøkelser, er sammenhengene i seg selv for svake, og det totale bildet for sammensatt og komplisert, til at det har noen hensikt å diskutere disse nærmere her.

5.4.2 Hva forteller de ikke-signifikante sammenhengene mellom de alternative konstruktene og elevenes resultater oss?

På tross av at det ikke finnes et solid grunnlag for å si noe ytterligere om sammenhengene som viste seg å være signifikante, kan det likevel være verdt å ta tak i alt det som *ikke* har vist seg å utgjøre en signifikant forskjell. Det første som bør nevnes her, er det faktum at EVALUERING, som var konstruktet der det i størst grad var en forskjell mellom Norge og Sverige, ikke viste seg å ha noen signifikant sammenheng med elevenes resultater. Det er likevel på sin plass å påpeke at dette ikke gir grunnlag for å trekke en tilsvarende konklusjon på vegne av instructional leadership som helhet, da diskusjonen over har vist at dette konstruktet kun måler et begrenset aspekt ved denne lederstilen. Konstruktets lave reliabilitet og de noe svakere ladninger fra itemene til konstruktet, bidrar selvfølgelig også til at det er vanskeligere å påvise slike signifikante sammenhenger.

Det samme argumentet kan derimot ikke brukes på det andre konstruktet der det ikke var noen signifikante sammenheng med elevenes resultater. Konstruktet MÅL, som både har god reliabilitet og der det jevnt over er sterke ladninger fra itemet mot konstruktet, kan nemlig heller ikke vise til noen signifikante sammenhenger med elevenes resultater. Dette skyldes altså neppe konstruktets psykometriske egenskaper, noe som gjør det mer nærliggende å tro at det rett og slett ikke er noen signifikant sammenheng å finne her. Som for de tidligere funnene, er det totale bildet likevel for komplekst til at det gir grunnlag å diskutere dette grundigere enn det som allerede er gjort her, ei heller konkludere med noe ut over dette.

Alt i alt er det mangelen på signifikante sammenhenger mellom konstruktene og elevenes resultater som er mest påfallende her, i det minste hvis man analyserer de norske og svenske dataene hver for seg. Dette kan henge sammen med flere faktorer. Først og fremst finnes det en rekke faktorer som disse modellene ikke tar høyde for, og som kan gjøre seg gjeldende på en rekke ulike måter. Det er ikke mulig å ta for seg alle disse, men ett slikt fenomen er allerede nevnt, nemlig at disiplinære problemer kan påvirke både konstruktet LÆRINGSMILJØ og elevenes resultater. Videre er spesielt Norge anerkjent som et land der svært lite av den totale variansen i elevresultater ligger på skolenivå, og følgelig også på rektornivå. Når denne studien i tillegg har vist at det er svakheter ved begrepsvaliditeten også hos de alternative konstruktene, er det ikke overraskende at det ikke er blitt gjort noen revolusjonerende funn her. Det skal sies at man ved å undersøke de norske og svenske dataene samlet fikk noe mer statistisk kraft, og således kunne påvise noen flere signifikante sammenhenger. Det totale bildet endret seg ikke nevneverdig ved å gjøre dette, og det kan tenkes at den statistiske styrken må forbedres ytterligere for at noe slikt skal påvises.

Selv om det altså ikke er gjort noen banebrytende oppdagelser i denne studien, er det kanskje nettopp dette som er bemerkelsesverdig; at rektors pedagogiske ledelse later til å ha en såpass neglisjerbar effekt på elevenes resultater. Funnene kan uansett fungere som et innspill i en diskusjon vedrørende om elementene ved norsk og svensk skole som reduserer rektors innflytelse som pedagogisk leder utgjør et pedagogisk problem for skolen? Trenger vi i det hele tatt rektorer? Før vi ser ytterligere på hva denne studien kan bidra med til forskningsfeltet, og hvor potensialet for videre studier eventuelt kan ligge, er det på sin plass å gå noe mer i dybden på hvilke utfordringer studien står overfor med tanke på dens validitet.

5.5 Validitet

Enhver studie må forholde seg til svake og sterke sider ved egen validitet, som ofte deles opp i indre og ytre validitet. Som nevnt, er den indre validiteten tilknyttet om studiens funn, og de analysene funnene bygger på, er egnet til å si noe meningsfullt om det studien er ment å belyse. I denne sammenheng blir jakten på alternative forklaringer, og svakheter ved det teoretiske og metodiske rammeverket som danner grunnlaget for funnene, av spesiell interesse. Den ytre validiteten dreier seg på sin side om hvorvidt studiens funn lar seg generalisere til populasjonen, som i dette tilfellet er norske og svenske 15-åringer og deres rektorer. Dette siste spørsmålet vil behandles først, da den ytre validiteten ikke er denne studiens største utfordring, men kanskje heller dens styrke. Den indre validiteten er derimot langt mer kompleks, og vil bli behandlet sist.

5.5.1 Hvordan er studiens ytre validitet?

OECD anses som en seriøs aktør innenfor storskalaundersøkelser, og det er lite grunnlag for å tro at det finnes aspekter ved trekningen av elevene, gjennomføringen av prøvene eller behandlingen av dataene som tilsier at resultatene fra PISA 2012 ikke skal kunne la seg generalisere på nasjonalt nivå. På tross av dette kan det fortsatt være på sin plass å bemerke at funn som tar utgangspunkt i rektorenes svar, ikke kan sies å være representative for landets rektorer på generell basis. Det er elevene i PISA 2012 som utgjør et representativt utvalg, ikke rektorene, som kun er representative i kraft av elevene sine. Denne nyansen i fortolkningen av resultatene har ingen reell betydning når rektorenes svar analyseres som variabler hos elevene, som i de strukturelle modellene, men kan derimot kan ha en begrenset effekt i de andre analysene. Hvordan funnene tilknyttet rektorene skal fortolkes, bør derfor ligge i bakhodet hvis resultatene fra denne studien analyseres eller brukes videre.

I tillegg bør det påpekes at en tverrsnittsundersøkelse som PISA 2012 må betraktes som et øyeblikksbilde av skolen, og kan verken si noe om situasjonen forut for eller i etterkant av undersøkelsen. Dette understreker også at studier basert på enkeltstående PISA-undersøkelser aldri vil kunne si noe om kausalitet eller andre forhold der utvikling over tid er en faktor.

5.5.2 Hvordan er studiens indre validitet?

En sentral del av denne studien er tilknyttet nettopp validitetsaspektet ved PISA 2012, men dreier seg altså ikke om studiens ytre validitet, men om begrepsvaliditeten til ledelseskonstruktene som benyttes i PISA, samt de alternative konstruktene. Som vist var det flere sider ved PISA-konstruktene som gjorde dem uegnet for videre analyser. Innsigelsene mot PISA-konstruktene og diskusjonen tilknyttet innholdsdimensjonen av begrepsvaliditeten til de alternative konstruktene vil i begrenset grad gjentas i denne delen, som heller vil fokusere på den delen av studiens indre validitet som er tilknyttet de statistiske analysene, altså konstruktens reliabilitet og modelltilpasningen i den konfirmerende faktoranalysen.

Det er flere styrker og svakheter knyttet til de statistiske aspektene ved de alternative konstruktene som allerede er bemerket, men som det kan være verdt å oppsummere. For det første er det primært Kaiser-kriteriet som, sammen med en teoretisk vurdering, er lagt til grunn for antall faktorer som beholdes i de eksplorerende faktoranalysene. Dette kan i noen tilfeller være for enkelt, selv om det her kan føyes til at også en scree plot-analyse⁴ viste at det kunne rettferdiggjøres å beholde fire faktorer, da plottet inneholdt to knekk, etter første og etter fjerde faktor. Selv om antall faktorer i de eksplorerende faktoranalysene med andre ord kan la seg forsvare, har det å skulle forene to ulike faktorløsninger i ett sett med alternative konstrukt vist seg å tvinge frem noen kompromisser som kan svekke validiteten til de berørte funnene. Spesielt i forbindelse med SC34Q05, ble de norske hensynene favorisert da itemet ble inkludert i konstruktet. Her var kjikvadratstestene nyttige, da de gjorde det mulig å avgjøre om modeller med ulikt antall frihetsgrader var signifikant forskjellige fra hverandre med tanke på modelltilpasning. Tilfellet over var det eneste tilfelle der en signifikant bedre svensk modell uten SC34Q05, riktignok kun på .05-nivå, ble forkastet til fordel for en signifikant bedre norsk modell på .01-nivå, som altså inkluderte SC34Q05. Et annet eksempel der kjikvadratstesten var nyttig, var i forbindelse med konstruktet EVALUERING, som hadde en så svak reliabilitet, spesielt med norske data, at det ikke hadde vært urimelig å forkaste dette. Ettersom det å beholde konstruktet likevel ikke gjorde modelltilpasningen signifikant dårligere, ble det teoretiske aspektet vektlagt, og konstruktet ble videreført. Dette valget kunne vært mer problematisk dersom konstruktet dannet grunnlaget for essensielle funn i studien, men slik var det altså ikke, og diskusjonene som gjøres på bakgrunn av frekvensanalysene lider ikke under at konstruktet som helhet er svakere enn ønskelig.

⁴ Se <http://ro.ecu.edu.au/cgi/viewcontent.cgi?article=1373&context=jephc> for ytterligere forklaring.

Den metodiske gjennomføringen inneholder også noen aspekter som kan ha betydning for studiens validitet, selv om det stort sett er redegjort for disse i metodedelen. For det første, kan det argumenteres for at man også i den eksplorerende fasen burde undersøkt rektorene i kraft av elevene de representerer, spesielt når det er en viss forskjell på fit-indikatorerne med de to utgangspunktene. Denne problematikken har riktignok vært kjent fra begynnelsen av, men har ikke blitt sett på som avgjørende, da samtlige konstrukter og funn ligger innenfor grenseverdiene til Hu og Bentler (1999), uavhengig av hvilket datasett som brukes i undersøkelsene, noe det også henvises til underveis. Det å kjøre de eksplorerende faktoranalyser på de flettede datafilene medfører også at utvalgsstørrelsen som analyseres later til å være på rundt 4700, når det reelle tallet rektorer ligger rundt 200, noe som også kan tenkes å gi et skjevt bilde. I tillegg til dette mulige validitetsproblemet, nevnes det i metodedelen også at det ikke lot seg gjøre å kombinere både replicate weights og plausible values i den strukturelle modellen. Det er vanskelig å anslå den reelle effekten av dette, men PISA sine egne analyser indikerer at det neppe utgjør en signifikant forskjell ved så store utvalg som her. Avslutningsvis kan det også nevnes at det kunne vært fordelaktig å utføre de konfirmerende faktoranalyser og de strukturelle modellene som tonivå-modeller der norske og svenske data undersøkes både separat og sammen, men dette lå på et høyere metodisk nivå enn det som var mulig å få gjennomført i denne omgang. Resultatene fra de strukturelle analysene er heller ikke kontrollert med tanke på andre skole- rektor- eller elevfaktorer som elevenes bakgrunn, som ble for omfattende i forbindelse med denne studien.

Den viktigste svakheten tilknyttet dataene dreier seg riktignok ikke om måten dataene behandles på, men at det er rektorene som selv rapporterer om sin egen pedagogiske ledelse. Dette kan spille en rolle gjennom at rektorene, både bevisst og ubevisst, modererer sine svar på bakgrunn av indre og ytre forhold eller en genuin misoppfatning av, i dette tilfellet, egen praksis. Selvrapporing er problematisk i forbindelse med all datainnsamling, kvantitativ som kvalitativ, og det er som regel umulig å anslå effekten av dette på en tilfredsstillende måte (Podsakoff og Organ, 1986). Bruken av flere item for å innhente informasjon om rektorenes pedagogiske ledelse og konfirmerende faktoranalyser kan dempe dette noe, men det hadde utvilsomt vært fordelaktig om PISA 2012, i likhet med TALIS-undersøkelsene, også hadde undersøkt et utvalg lærere ved hver skole, da slike undersøkelser kan bidra til å moderere og utdype eller potensielt underbygge rektorenes svar (Atwater, Waldman, og Brett, 2002, s. 196). Det knytter seg også noe usikkerhet tilknyttet at skole spørreskjemaet kun spør hva rektor gjør, og således ikke kartlegger hva andre personer i ledelsen håndterer.

Som nevnt knytter det seg også noe usikkerhet til bruken av Likert-skalaer i forbindelse med parametriske analyser, og selv om skalaen tilknyttet SC34 holder seg innenfor det som er det anbefalte antall svarkategorier, og itemene for det meste behandles gruppevis, kunne dette like fullt vært sikret bedre ved å bruke Item Response Theory (ITR) for å skalere konstruktene, noe som gjøres i PISA (OECD, 2012), men som ble for omfattende her. Derimot er det en styrke at rektorene i de to landene deler stort sett er like på gruppenivå med tanke på jobbakgrunn, alders- og kjønnsammensetning, som ellers kunne bidratt til å gjøre det vanskeligere å sammenligne de to gruppene, da eventuelle forskjeller nå ikke kan avskrives på bakgrunn av at rektorgruppene er for ulike.

5.6 Studiens bidrag til forskningsfeltet

I innledningen kan man lese at denne studien, gjennom å analysere begrepsvaliditeten til ledelseskonstruktene i PISA 2012, skal bidra til å styrke forskningsfeltet i nordisk sammenheng med tanke på fortolkningen av både eksisterende og fremtidige data. I denne delen vil forskningsspørsmålene danne utgangspunkt for en kort gjennomgang av studiens funn, deres bidrag til forskningsfeltet og mulige områder for videre forskning.

5.6.1 Begrepsvaliditeten til PISA-konstruktene

Funnene tilknyttet begrepsvalideringen av konstruktene tilknyttet PISA sin Index of Principal Management, viser at modelltilpasningen til indeksen som helhet ikke er god nok til at indeksen kan benyttes i videre strukturelle analyser for Norges og Sveriges del. Dette kan virke overraskende, spesielt med tanke på at Cronbachs α stort sett er akseptable, og det er uheldig at PISA-publikasjonene som bruker indeksen og de tilhørende konstruktene, ikke redegjør bedre for det teoretiske eller empiriske grunnlaget som konstruktene bygger på. I de fremtidige PISA-undersøkelsene kan det være fordelaktig om antall item innen hvert konstrukt økes til minst fire, slik at man kan bruke konfirmerende faktoranalyser på konstruktene hver for seg, slik det var mulig å gjøre med LEADCOM her. Dette gir en mulighet for å etterprøve resultatene fra reliabilitetsanalysene med et mer pålitelig analytisk verktøy. Den innholdsmessige delen av begrepsvalideringen tyder også på at flere av itemene, så vel som hovedspørsmålet som helhet, fortsatt har et utviklingspotensial. Hvis det fremover skal være et mål å kunne benytte fremtidige PISA-undersøkelser til å undersøke sammenhengen mellom rektors pedagogiske ledelse på en tilfredsstillende måte, ville det

utvilsomt vært en fordel om hovedspørsmålet tilknyttet pedagogisk ledelse ble betydelig utvidet. Dersom det lar seg gjøre å få etablert gode konstruktører som lot seg bruke over tid, kunne PISA-undersøkelsene ikke bare sporet endringer i elevenes prestasjoner over tid, men også endringer i rektors ledelse. Erfaringene fra denne studien viser dessuten at det hadde vært fordelaktig om en større del av det teoretiske rammeverk som ligger til bakgrunn for de ulike itemene og konstruktene var offentlig tilgjengelige, slik at de lettere lot seg etterprøve av utenforstående.

5.6.2 Begrepsvaliditeten til de alternative konstruktene

På tross av en uakseptabel modelltilpasning for PISA-konstruktene, viser denne studien at det likevel kan ligge et analytisk potensiale i itemene tilknyttet rektors pedagogiske ledelse, selv om diskusjonen over viser at det finnes flere problematiske aspekter ved innholdsdimensjonen til de alternative konstruktene også. For å styrke holdbarheten til konstruktene ytterligere, ville verdien av et lærerspørreskjema, som vi har i TALIS, vært stor, da dette kan fungere som en kontrollmekanisme i forhold til rektorenes selvrapporing. Modelltilpasningen viste seg i det minste å bli akseptabel for begge landene til slutt, og de alternative konstruktene kan potensielt sett være et innlegg i en debatt rundt videreutviklingen av dette hovedspørsmålet i fremtidige skolespørreskjemaer.

5.6.3 Likheter og ulikheter mellom Norge og Sverige

Denne delen av studien viste at det finnes flede nevneverdige likheter så vel som bemerkelsesverdige forskjeller mellom rektorene i de to landene, men likhetene og ulikhetene kunne muligens fremstått tydeligere i en tilsvarende studie med flere land, da disse kan utgjøre et mer omfattende bakteppe for sammenligningene. Analysene i denne studien viste at målstyringsprinsipper er fremtredende i begge land, samt at rektorene bruker mye tid på å støtte lærerne når det oppstår problemer i klasserommet. Svenske rektorer later like fullt å være noe mer farget av ledelsesfunksjoner som forbindes med instructional leadership, blant annet tilknyttet arbeidet med skolens mål, rektorinitiert håndtering av klasseromsproblematikk og et mer utbredt fokus på bruk av observasjon og evaluering av undervisningspraksisen ved skolen. Bakgrunnen for dette er interessant, og det kunne i denne sammenheng vært spennende med nye studier som undersøker dette. Dette kan eksempelvis gjøres gjennom en komparativ analyse av innholdet i rektorutdannelsen med fokus på tilstedeværelsen av ulike

lederstiler, men det kunne også vært spennende å se på rektorenes bevissthet rundt egen praksis tilknyttet instructional leadership og andre relevante lederstiler, kanskje spesielt i Sverige. Også i forbindelse med sammenligninger på tvers av landegrensener vil det være en fordel om det ble etablert ledelseskonstrukt som kunne stå uendret i flere undersøkelser.

5.6.4 Rektors pedagogiske ledelse og elevenes resultater

Selv om resultatene av de strukturelle modellene ikke inneholdt noen revolusjonerende funn, kan fraværet av dette ses på som et funn i seg selv. Dette kan selvfølgelig henge sammen med at studiens design og statistiske styrke ikke er tilstrekkelig for å avdekke eventuelle effekter av rektors ledelse, men kan også ligge i det substansielle. Er mekanismene i norsk og svensk skole som innskrenker rektorenes handlingsrom så omfattende at de slår ihjel alle utviklingstiltak? Er kunnskapen og viljen som skal til for å utvikle og forbedre undervisningen ikke til stede i stor nok grad? Eller er det rett og slett slik at skoleledelse har en neglisjerbar effekt på elevenes læringsutbytte? I tillegg til å være ett innlegg i denne debatten, finnes det utvilsomt et behov for å gå disse resultatene etter i sømmene, og ikke minst å kontrollere for ulike skole- og elevfaktorer, som elevenes bakgrunn eller disiplinære problemer ved skolen, noe som ikke er gjort i denne studien. Nye modeller, basert på norske og svenske data med utgangspunkt i mer avanserte flernivåanalyser, er selvfølgelig også en fordel å få gjennomført, noe som ikke er gjort her. Mulighetene for å utvikle strukturelle modeller som undersøker sammenhengene mellom ulike skolefaktorer og rektors pedagogiske ledelse er også til stede, som hvilken sammenheng det er mellom skolestørrelse og de alternative konstruktene, forskjeller mellom by og land, kjønn, eller for Sveriges del mellom de private og de offentlige skolene.

Det analytiske potensialet til storskalaundersøkelser som PISA er nærmest ubegrenset, i det minste innenfor den kvantitative delen av utdanningsforskningen. PISA sitt longitudinale perspektiv kan også vise seg å bli verdifullt her, spesielt hvis det lar seg gjøre å etablere gode ledelseskonstrukt som kan benyttes over tid. Dette vil gjøre det mulig å spore effekten av rektors ledelse over tid, som igjen kan bidra til å gi analysene langt mer dybde og funnene mer tyngde. Gjennom å øke fokuset på og kvaliteten til spørreskjemaet som skal belyse rektors ledelse, vil vi kanskje se at det er skoleledelse som preger avisoverskriftene når rapporten fra PISA 2021 slippes, og ikke elevenes plass på de internasjonale rankinglistene.

6 LITTERATUR

- Aasen, P., Møller, J., Rye, E., Ottesen, E., Prøitz, T. S., & Hertzberg, F. (2012). Kunnskapsløftet som styringsreform—et løft eller et løfte? *Forvaltningsnivåenes og institusjonenes rolle i implementeringen av reformen*.
- Atwater, L. E., Waldman, D. A., & Brett, J. F. (2002). Understanding and optimizing multisource feedback. *Human Resource Management, 41*(2), 193-208.
- Bass, B. M. (1997). Does the transactional—transformational leadership paradigm transcend organizational and national boundaries? *American Psychologist, 52*(2), 130-9.
- Berg, G. (1990). *Skolledning och professionellt skolledarskap* (Vol. 92). Uppsala: Uppsala Universit t.
- Blase, J. R., & Blase, J. (1998). *Handbook of instructional leadership: How really good principals promote teaching and learning*: Corwin Press Thousand Oaks, CA.
- Bolman, L. G., & Deal, T. E. (2005). *Nytt perspektiv p  organisasjon og ledelse: strukturer, sosiale relasjoner, politikk og symboler* (3. utgave ed.). Oslo: Gyldendal akademisk.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.
- Bryman, A., & Cramer, D. (2011). *Quantitative data analysis with IBM SPSS 17, 18 and 19: a guide for social scientists*. London: Routledge.
- Byrne, B. M. (2012). *Structural equation modeling with Mplus: basic concepts, applications, and programming*. New York: Routledge.
- Carifio, J., & Perla, R. J. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *Journal of Social Sciences, 3*(3), 106.
- Carlsten, T. C., Caspersen, J., Vibe, N., & Aamodt, P. O. (2014). Resultater fra TALIS 2013: Norske funn fra ungdomstrinnet i internasjonalt lys. Oslo: NIFU Nordisk institutt for studier av innovasjon, forskning og utdanning.
- Caspersen, J. (2011). Skolekultur og elevresultater - Hvilke muligheter gir TALIS-unders kelsen? Oslo: NIFU.
- Cohen, L., Manion, L., Morrison, K., & Bell, R. C. (2011). *Research methods in education*. London: Routledge.
- Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: theory and application. *The American journal of medicine, 119*(2), 166. e7-. e16.
- Costello, A. B., & Osborne, J. W. (2005). Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. *Practical Assessment Research & Evaluation, 10*(7), 2.

- Davey, A. (2005). Issues in Evaluating Model Fit With Missing Data. *Structural Equation Modeling: A Multidisciplinary Journal*, 12(4), 578-97.
- Emstad, A. B., & Postholm, M. B. (2010). "Instructional leadership" – et godt utgangspunkt for ledelse av skolens læringsarbeid. I R.-A. Andreassen, E. J. Irgens & E. M. Skaalvik (Red.), *Kompetent skoleledelse* (s. 183-95). Trondheim: Tapir akademisk forl.
- Engeland, Ø., & Langfeldt, G. (2009). Forholdet mellom stat og kommune i styring av norsk utdanningspolitikk 1970-2008. *Acta Didactica Norge*, 3(1), (Art. 9, 16 sider).
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics: and sex and drugs and rock 'n' roll*. Los Angeles: SAGE.
- Forsberg, E. (2010). Styring av den obligatoriska skolan - mellan stabilitet och förändring. I U. P. Lundgren, R. Säljö & C. Liberg (Red.), *Lärande, skola, bildning* (s. 451 - 74). Stockholm: Natur & kultur.
- Fredriksson, U., Oskarsson, M., Pettersson, A., Karlsson, K., Sollerman, S., Eliasson, N., . . . Sundgren, M. (2013). PISA 2012: 15-åringars kunskaper i matematik, läsförståelse och naturvetenskap.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual review of psychology*, 60, 549-76.
- Hallerström, H. (2006). *Rektorers normer i ledarskapet för skolutveckling*. Sociologiska institutionen, Lunds universitet, Lund.
- Hallinger, P. (1994). A resource manual for the principal instructional management rating scale. *Nashville [TN]: Vanderbilt University, Centre for advanced study of educational leadership*.
- Hallinger, P. (2003). Leading educational change: Reflections on the practice of instructional and transformational leadership. *Cambridge Journal of education*, 33(3), 329-52.
- Hallinger, P. (2011). Leadership for learning: Lessons from 40 years of empirical research. *Journal of Educational Administration*, 49(2), 125-42.
- Hallinger, P., & Heck, R. H. (1998). Exploring the Principal's Contribution to School Effectiveness: 1980-1995*. *School effectiveness and school improvement*, 9(2), 157-91.
- Hallinger, P., & Heck, R. H. (2010). Collaborative leadership and school improvement: Understanding the impact on school capacity and student learning. *School leadership and management*, 30(2), 95-110.
- Hallinger, P., & Murphy, J. F. (1985). Assessing the instructional management behavior of principals. *The Elementary School Journal*.
- Hansen, K. Y., Gustafsson, J.-E., Rosén, M., Sulkunen, S., Nissinen, K., Kupari, P., . . . Rønberg, L. (2014). *Northern Lights on TIMSS and PIRLS 2011: Differences and similarities in the Nordic countries*: Nordic Council of Ministers.

- Haug, B. (2010). Desentralisering og skoleprestasjoner. I K. Sivesind & E. Elstad (Red.), *PISA: sannheten om skolen?* (s. 246-57). Oslo: Universitetsforl.
- Helgøy, I. (2006). Rhetoric and Action in Regulating the Public Schools in Norway and Sweden. *Scandinavian Political Studies*, 29(2), 89-110.
- Helgøy, I., & Homme, A. (2004). *Governance in primary and lower secondary education: comparing Norway, Sweden and England* (Vol. 16-2004). Bergen: Rokkansenteret.
- Helgøy, I., & Homme, A. (2006). Policy tools and institutional change: Comparing education policies in Norway, Sweden and England. *Journal of Public Policy*, 26(02), 141-65.
- Helgøy, I., & Homme, A. (2007). Towards a new professionalism in school? A comparative study of teacher autonomy in Norway and Sweden. *European Educational Research Journal*, 6(3), 232-49.
- Hu, L. t., & 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.
- Hybertsen Lysø, I., Stenaker, B., Aamodt, P. O., & Mjøen, K. (2011). Ledet til ledelse: nasjonal rektorutdanning i grunn- og videregående skole i et internasjonalt perspektiv. Delrapport 1 *Evaluering av den norske rektorutdanningen*. Oslo: NIFU.
- Hybertsen Lysø, I., Stensaker, B., Røthe, R., Federici, R., Schei Olsen, M., & Solem, A. (2012). Ledet til lederutvikling. Nasjonal rektorutdanning i grunn- og videregående skole; forskjeller og likheter mellom de seks programtilbudene. Delrapport 2 *Evaluering av den nasjonale rektorutdanningen*. Oslo: NIFU.
- Imsen, G. (2004). Skolens ledelse, skolens kultur og praksis i klasserommet: Er det noen sammenheng? I G. Imsen (Red.), *Det ustyrige klasserommet: om styring, samarbeid og læringsmiljø i grunnskolen* (s. 144 - 64). Oslo: Universitetsforl.
- Imsen, G., & Volckmar, N. (2014). The Norwegian School for All: Historical Emergence and Neoliberal Confrontation. I U. Blossing, G. Imsen & L. Moos (Red.), *The Nordic Education Model* (Vol. 1, s. 35-55): Springer Netherlands.
- Jacobson, S. L., & Ylimaki, R. M. (2011). Comparative perspectives: an overview of seven educational contexts *US and Cross-National Policies, Practices, and Preparation* (s. 1-16): Springer.
- Johansson, O., & Bredeson, P. V. (2011). Framtida forskningsperspektiv på rektor - Vilken forskning saknas? I O. Johansson (Red.), *Rektor - En Forskningsöversikt 2000-2010* (s. 61-74). Stockholm: Vetenskapsrådet.
- Kjærnsli, M., & Jensen, F. (2013). Læringsmiljøet i skolen. I M. Kjærnsli & R. V. Olsen (Red.), *Fortsatt en vei å gå - Norske elevers kompetanse i matematikk, naturfag og lesing i PISA 2012* (s. 261-75). Oslo: Univeristetsforlaget.

- Kjærnsli, M., & Olsen, R. V. (2013a). PISA 2012 - sentrale funn. I M. Kjærnsli & R. V. Olsen (Red.), *Fortsatt en vei å gå - Norske elevers kompetanse i matematikk, naturfag og lesing i PISA 2012* (s. 13-42). Oslo: Univeristetsforlaget.
- Kjærnsli, M., & Olsen, R. V. (Red.). (2013b). *Fortsatt en vei å gå - Norske elevers kompetanse i matematikk, naturfag og lesing i PISA 2012*. Oslo: Univeristetsforlaget.
- Kleven, T. A. (2002). Ikke-eksperimentelle design. I T. Lund (Red.), *Innføring i forskningsmetodologi* (s. 265-86). Oslo: Unipub.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling*. New York: Guilford Press.
- Kuzon, W. M., Urbanek, M. G., & McCabe, S. (1996). The seven deadly sins of statistical analysis. *Annals of plastic surgery*, 37(3), 265-72.
- Langfeldt, G. (2008). *Ansvar og kvalitet: strategier for styring i skolen*. Oslo: Cappelen akademisk.
- Leithwood, K., & Jantzi, D. (2000). The effects of transformational leadership on organizational conditions and student engagement with school. *Journal of Educational Administration*, 38(2), 112-29.
- Leithwood, K., Jantzi, D., & Steinbach, R. (1999). *Changing leadership for changing times*: McGraw-Hill International.
- Lillejord, S. (2011). Kunsten å være rektor. I J. Møller & E. Ottesen (Red.), *Rektor som leder og sjef: om styring, ledelse og kunnskapsutvikling i skolen* (s. 284-301). Oslo: Universitetsforlaget.
- Lundahl, L. (2002). Sweden: decentralization, deregulation, quasi-markets - and then what? *Journal of Education Policy*, 17(6), 687-97.
- Lundahl, L., Erixon Arreman, I., Holm, A.-S., & Lundström, U. (2013). Educational marketization the Swedish way. *Education Inquiry*, 4(3).
- Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In Search of Golden Rules: Comment on Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler's (1999) Findings. *Structural Equation Modeling: A Multidisciplinary Journal*, 11(3), 320-41.
- Miller, L. A., McIntire, S. A., & Lovler, R. L. (2011). *Foundations of psychological testing: a practical approach* (3rd ed. ed.). Los Angeles: Sage Publishing.
- Møller, J. (2006a). Hvilke svar gir forskning om god skoleledelse? I J. Møller & O. L. Fuglestad (Red.), *Ledelse i anerkjente skoler* (s. 27-44). Oslo: Universitetsforlaget.
- Møller, J. (2006b). Nyere forskning om skole-ledelse i gode skoler. *Norsk pedagogisk tidsskrift*, 90(02).
- Møller, J. (2011). Rektorers profesjonsforståelse - faglig autonomi og administrativ underordning. I J. Møller & E. Ottesen (Red.), *Rektor som leder og sjef: om styring, ledelse og kunnskapsutvikling i skolen* (s. 27-50). Oslo: Universitetsforlaget.

- Møller, J., Ottesen, E., Dyrdal Solbrekke, T., & Aaserud, T. (2000). Kartlegging og analyse av rektors arbeidsvilkår i Norge. I L. Moos (Red.), *Skoleledelse i Norden: en kortlægning af skoleledernes arbeidsvilkår, rammebetingelser og opgaver : en rapport til Nordisk Ministerråd* (s. 197 - 257). København: Nordisk Ministerråd.
- Møller, J., & Skedsmo, G. (2013). Modernising education: New Public Management reform in the Norwegian education system. *Journal of Educational Administration and History*, 45(4), 336-53.
- Nihlfors, E. (2010). Skolledare - att förverkliga det nationella uppdraget. I U. P. Lundgren, R. Säljö & C. Liberg (Red.), *Lärande, skola, bildning* (s. 493 - 510). Stockholm: Natur & kultur.
- Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. *Advances in health sciences education*, 15(5), 625-32.
- NOU 2002:10. (2002). *Førsteklasses fra første klasse. Forslag til rammeverk for et nasjonalt kvalitetsvurderingssystem av norsk grunnopplæring*. Oslo: Utdannings- og forskningsdepartementet.
- Nygren, A. M., & Johansson, O. (2000). Den Svenske rektorn efter 1945. I L. Moos (Red.), *Skoleledelse i Norden: en kortlægning af skoleledernes arbeidsvilkår, rammebetingelser og opgaver : en rapport til Nordisk Ministerråd* (s. 258 - 311). København: Nordisk Ministerråd.
- OECD. (2002). Education at a glance 2002: OECD publishing.
- OECD. (2008a). Education at a glance 2008: OECD Indicators: OECD publishing.
- OECD. (2008b). *Improving School Leadership, Volume 1 Policy and Practice: Policy and Practice* (Vol. 1): OECD publishing.
- OECD. (2008c). *Improving School Leadership, Volume 2: Case Studies on System Leadership* (Vol. 2): OECD Publishing.
- OECD. (2009). *PISA Data Analysis Manual: SPSS, Second Edition* (2. utgave ed.): OECD Publishing.
- OECD. (2010a). *PISA 2009 Results: What Makes a School Successful?* : OECD Publishing.
- OECD. (2010b). *TALIS 2008 Technical Report*: OECD Publishing.
- OECD. (2011). Quality Framework and Guidelines For OECD Statistical Activities. (Version 2011/1).
- OECD. (2012). *PISA 2009 Technical Report*: OECD Publishing.
- OECD. (2013a). *PISA 2012 Assessment and Analytical Framework*: OECD Publishing.
- OECD. (2013b). *PISA 2012 Results: Excellence Through Equity (Volume II): Preliminary Version - Giving Every Student the Chance to Succeed*: OECD Publishing.
- OECD. (2013c). PISA In Focus #32: Do students perform better in schools with orderly classrooms? *PISA In Focus* (pp. 4): OECD publishing.

- OECD. (2013d). Synergies for better learning: An international perspective on evaluation and assessment. *OECD Reviews of Evaluation and Assessment in Education*.
- OECD. (2014a). Education at a glance 2014: OECD Indicators: OECD publishing
- OECD. (2014b). *TALIS 2013 Results: An International Perspective on Teaching and Learning*: OECD Publishing.
- Pierre, J. (2007). *Skolan som politisk organisation*. Malmö: Gleerups utbildning AB.
- Podsakoff, P. M., & Organ, D. W. (1986). Self-Reports in Organizational Research: Problems and Prospects. *Journal of Management*, 12(4), 531-44.
- Richardson, G. (2004). *Svensk utbildningshistoria: skola och samhälle förr och nu* (7. rev. uppl. ed.). Lund: Studentlitteratur.
- Robinson, V. M. (2010). From instructional leadership to leadership capabilities: Empirical findings and methodological challenges. *Leadership and Policy in Schools*, 9(1), 1-26.
- Robinson, V. M. J., Lloyd, C. A., & Rowe, K. J. (2008). The Impact of Leadership on Student Outcomes: An Analysis of the Differential Effects of Leadership Types. *Educational Administration Quarterly*, 44(5), 635-74.
- Roe, A., & Kjærnsli, M. (2013). Resultater fra skole spørreskjemaet. I M. Kjærnsli & R. V. Olsen (Red.), *Fortsatt en vei å gå - Norske elevers kompetanse i matematikk, naturfag og lesing i PISA 2012* (s. 237-60): Univeristetsforlaget.
- Sejersted, F. (2005). *Sosialdemokratiets tidsalder: Norge og Sverige i det 20. århundre*. Oslo: Pax.
- Sergiovanni, T. J. (2009). *The principalship: a reflective practice perspective*. Boston: Pearson.
- Sivesind, K., & Elstad, E. (2010). *PISA: sannheten om skolen?* Oslo: Universitetsforl.
- Skedsmo, G. (2011). Vurdering som styring av utvikling og overvåkning av resultater. I J. Møller & E. Ottesen (Red.), *Rektor som leder og sjef: om styring, ledelse og kunnskapsutvikling i skolen* (s. 74-94). Oslo: Universitetsforlaget.
- Skolverket. (1998). *Nationella kvalitetsgranskningar 1998, Skolverkets rapport nr 160*. Stockholm: Skolverket.
- SOU. (2014:5). *Staten får inte abdikera - om kommunalisering*. Stockholm: Utbildningsdepartementet.
- Southworth, G. (2002). Instructional Leadership in Schools: Reflections and empirical evidence. *School Leadership & Management*, 22(1), 73-91.
- St.meld. nr. 30. (2003-2004). *Kultur for læring*. Oslo: Stortinget
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis: understanding concepts and applications*. Washington, D.C.: American Psychological Association.

- Thuen, H. (2010). Skolen – et liberalistisk prosjekt? 1860–2010. *Norsk pedagogisk tidsskrift*, 94 E(04).
- Ullman, A. (1997). *Rektorn: en studie av en titel och dess bärare*. HLS Förlag, Stockholm.
- Von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful. *IERI monograph series*, 2, 9-36.
- Wang, J., & Wang, X. (2012). *Structural equation modeling: applications using Mplus*. Chichester: Wiley.
- Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, 31(2–3), 114-28.

7 VEDLEGG

7.1 Item tilknyttet pedagogisk ledelse i PISA 2012


ITEM	(item som også finnes i PISA 2009 står i kursiv)
SC34Q01	Jeg arbeider for å fremme skolens omdømme i lokalsamfunnet.
SC34Q02	<i>Jeg bruker elevens faglige resultater for å utvikle skolens pedagogiske mål.</i>
SC34Q03	<i>Jeg forsikrer meg om at lærernes utvikling er i samsvar med skolens pedagogiske mål.</i>
SC34Q04	Jeg sørger for at lærernes arbeid er i samsvar med skolens pedagogiske mål.
SC34Q05	Jeg oppfordrer til undervisningspraksis basert på nyere forskning.
SC34Q06	Jeg roser lærere som har elever som deltar aktivt i undervisningen.
SC34Q07	<i>Jeg tar initiativ til å diskutere årsakene når en lærer har problemer i klasserommet sitt.</i>
SC34Q08	Jeg gjør lærere oppmerksomme på viktigheten av å utvikle elevenes kritiske og sosiale evner.
SC34Q09	<i>Jeg er opptatt av forstyrrende atferd i klasserommene.</i>
SC34Q10	Jeg gir ansatte muligheten til å delta i beslutningsprosesser angående skolen.
SC34Q11	Jeg motiverer lærere til å utvikle en skolekultur basert på stadige forbedringer.
SC34Q12	Jeg ber lærere delta i evalueringen av ledelsens arbeid.
SC34Q13	<i>Når en lærer kommer med et problem knyttet til arbeidet i klasserommet, løser vi det sammen.</i>
SC34Q14	Jeg diskuterer skolens faglige mål med lærere på fellesmøter.
SC34Q15	Jeg henviser til skolens faglige mål når lærere deltar i avgjørelser angående fagplaner.
SC34Q16	Jeg diskuterer læringsresultater med lærerne for å kartlegge styrker og svakheter i skolens fagplaner.
SC34Q17	Jeg leder eller deltar i etterutdanningsaktiviteter med lærerne.
SC34Q18	Jeg setter av tid på felles møter slik at lærere kan dele ideer og utveksle erfaringer fra etterutdanningsaktiviteter.
SC34Q19	Jeg gjennomfører uformelle observasjoner i klasserommet jevnlig, (uformelle observasjoner er vanligvis ikke avtalt, varer minst fem minutter, og kan, men må ikke, inneholde skriftlig tilbakemelding eller et formelt møte).
SC34Q20	Jeg vurderer skriftlige elevarbeider når jeg evaluerer undervisningen.
SC34Q22	Jeg vurderer lærernes innsats.

Table 6.1

School principal leadership behavioral items (2007-08)
Items in behavioral indices for principals in lower secondary education

Indices	Item name	Description of the item
Management-school goals index	bcg15a	I make sure that the professional development activities of teachers are in accordance with the teaching goals of the school.
	bcg15b	I ensure that teachers work according to the school's educational goals.
	bcg15d	I use student performance results to develop the school's educational goals.
	bcg15j	I take exam results into account in decisions regarding curriculum development.
	bcg15k	I ensure that there is clarity concerning the responsibility for co-ordinating the curriculum.
	bcg16m	In this school, we work on goals and/or a school development plan.
Instructional management index	bcg15g	When a teacher has problems in his/her classroom, I take the initiative to discuss matters.
	bcg15h	I inform teachers about possibilities for updating their knowledge and skills.
	bcg15l	When a teacher brings up a classroom problem, we solve the problem together.
	bcg15m	I pay attention to disruptive behaviour in classrooms.
Direct supervision of instruction in the school index	bcg15c	I observe instruction in classrooms.
	bcg15e	I give teachers suggestions as to how they can improve their teaching.
	bcg15f	I monitor students' work.
Accountable management index	bcg15i	I check to see whether classroom activities are in keeping with our educational goals.
	bcg16a	An important part of my job is to ensure ministry approved instructional approaches are explained to new teachers, and that more experienced teachers are using these approaches.
	bcg16d	A main part of my job is to ensure that the teaching skills of the staff are always improving.
	bcg16e	An important part of my job is to ensure that teachers are held accountable for the attainment of the school's goals.
	bcg16f	An important part of my job is to present new ideas to the parents in a convincing way.
	bcg16h	It is important for the school that I see to it that everyone sticks to the rules.
Bureaucratic management index	bcg16i	It is important for the school that I check for mistakes and errors in administrative procedures and reports.
	bcg16j	An important part of my job is to resolve problems with the timetable and/or lesson planning.
	bcg16k	An important part of my job is to create an orderly atmosphere in the school.
	bcg16o	I stimulate a task-oriented atmosphere in this school.

Source: OECD, TALIS Database.

StatLink  <http://dx.doi.org/10.1787/608025205225>

7.2 Item tilknyttet “School management” i TALIS 2008