

Sammendrag:

Tester av Stated Preference-teknikker og samvalgdesign

En metodestudie av kollektivtrafikanter verdsetting av kvalitetsfaktorer

Dette prosjektet har hatt som formål å kritisk gjennomgå design av *Stated Preference* (SP) undersøkelser for å avdekke måter designet påvirker resultatet og gi anbefalinger til hvordan problemet kan reduseres.

Hovedproblemstillingen har vært todelt:

1. I hvilken grad er det mulig å utvikle tester som avdekker hva som er god eller dårlig design på SP-undersøkelser, og kan disse testene benyttes for å vurdere hvilke design som fungerer best?
2. Vil egenadministrerte PC-intervjuer fungere like bra som hjemmeintervjuer?

Det er blitt gjennomført en pilotstudie blant Oslos kollektivreisende, der 10 ulike design for undersøkelsen er blitt testet og analysert. Med *design* forstår vi:

- Utvalgs- og rekrutteringsmetode
- Metode for datainnsamling
- Struktur på spørreskjema
- SP-metoder
- Utforming av de enkelte SP-metodene

Undersøkelsen omfattet fire SP-metoder: Parvise samvalg (SC), likeverdpris (CV/TP), Frisch-metoden (FM) og Fleksibelt pakkevalg (POG).

I SC verdsettes flere goder i forhold til hverandre, og disse godene kan både være markedsgoder og ikke-markedsgoder. I undersøkelsene blir respondenter stilt overfor valg mellom reisealternativer som består av en "pakke" av egenskaper (f.eks. reisetid, pris, regularitet og komfort) ved en bestemt reise. I undersøkelsen inngikk 6 ulike SC-design.

CV/TP er mer direkte spørsmål om betalingsvillighet av typen "*Hva er det maksimale du ville betalt for reisen med økt reisetid?*"

FM er en metode som sirkler inn en respondents avveining mellom to og to egenskaper. I en serie valg

mellom to reisealternativer blir forskjellene i nivåene på egenskapene gradvis redusert i retning av respondentens preferanser.

POG lar respondenten velge den kombinasjonen av godeforbedringer som gir størst nytte. Samtidig må respondenten ta hensyn til at dette påvirker prisen.

Datainnsamlingsmetode og frafall

Datainnsamlingen har vært fordelt mellom "tradisjonelle" hjemmeintervjuer og egenadministrerte, PC-baserte intervjuer på diskett som respondentene har fått tilsendt.

Det er en viss forskjell mellom de to utvalgene, som i noen grad skyldes at forskjellige grupper har tilgang til PC og vegrer seg mot hjemmeintervju. I et hovedprosjekt er det derfor nødvendig å supplere egenadministrerte intervjuer med hjemmeintervjuer blant underrepresenterte respondentgrupper.

I telefonrekrutteringen av respondenter var om lag 12 % villige til å delta i egenadministrerte intervjuer, og oppfylte kriteriene for å delta i undersøkelsen. Til sammenligning var andelen 7 % for rekrutteringen til hjemmeintervjuer.

Frafallet blant dem som fikk tilsendt diskett for egenadministrerte intervjuer ligger jevnt på ca 50 %. Andelen kan reduseres ved å utvikle mer brukervennlige løsninger som ikke avhenger av operativsystem, f.eks. internett-baserte løsninger.

Tester av designet

I designtestene er hovedformålet å påvise måter designet har påvirket respondentenes svar, og måter dette problemet kan reduseres. Vi har sett på forhold ved designet som påvirker

- størrelse, signifikans og fortegn på verdsettingene.

- leksikografi, dvs når respondenten sorterer alle alternativene i SC etter én (eller flere) egenskaper gjennom hele samvalgsekvensen. Leksikografi er et alvorlig metodeproblem i samvalganalyser, og kan skyldes manglende balansering av egenskapsnivåene (ev. spesielle preferanser), eller at respondenten forenkler beslutningsrutinen. Det er også en viss sannsynlighet for at det skjer tilfeldig.
- forekomsten av avvikende svar, dvs. svar det ut fra modellene er svært liten (i denne studien definert som hhv. 0,05 og 0,1 prosent) sjans for å velge.
- 0-svar og overbud: 0-svar omfatter 0 i verdsetting og 0 i betalingsvilje, mens overbud innebærer ekstremt høy verdsetting av en kvalitetsfaktor.
- tretthetstegn, som fører til mindre konsistent svargivning.
- konsistens i svargivningen på tvers av SP-metodene
- respondentenes vurdering av vanskelighetsgrad ved metodene.
- frafall.

Samvalgene

I alt seks ulike samvalgsdesign er blitt utarbeidet:

1. Standarddesign: Fire kvalitetsegenskaper med tre nivåer hver og opptil 9 valg. Preferanseusikkerhet (kan velge "*Absolutt alternativ X*" eller "*Trolig alternativ X*").
2. Balansering av nivåene ved hjelp av CV/TP.
3. CV/TP-balansering, samt svaralternativ "Vet ikke" og "Like gode".
4. Standarddesign, og i tillegg et eget blokkdesign med grupper av faktorer på flere nivåer.
5. CV/TP-balansering. Ta med opptil 15 valg for å studere tretthetseffekter.
6. Redusere antall faktorer fra 4 til 3 for å se hvordan dette påvirker leksikografi og verdsetting.

I samvalgsekvensene er det til dels store sprik i verdsettingen av kvalitetsfaktorer mellom designvariantene. Tabell S.1 viser et utdrag av beregningene. Noe av forskjellene skyldes tilfeldig variasjon i de små utvalgene, men tester viser at forskjellene også skyldes egenskaper ved designet.

- Bedre balansering av nivåene i SC-designet gir lavere andeler leksikografi og forbedrer samvalgsekvensene betraktelig. Med bakgrunn i dette er det forsøkt å forbedre balanseringen ved å la variasjonen rundt basisprisen bli bestemt av respondentenes svar i en forutgående CV/TP-sekvens. Dette gir lavere andeler leksikografi. Et slikt opplegg med å la SP-metodene spille sammen har med andre ord hjulpet til med å redusere et viktig metodeproblem.
- Samvalgsekvenser med 15 valgoganger oppfattes av respondentene som vanskeligere å besvare enn sekvenser med opptil 9 valgoganger. De har fungert relativt dårlig med hensyn til de vanligste indikatorene for forklaringskraft, og har høyere andeler avvikende svar. Lange samvalgsekvenser skaper derfor usikkerhet rundt beregningene snarere enn å gi oss mer informasjon. Våre analyser indikerer videre at lange samvalgsekvenser ikke bare påvirker den enkelte samvalgsekvensen, men også senere sekvenser i samme intervju.
- Ved å åpne for preferanseusikkerhet i svargivningen kan vi hente mer informasjon fra hver observasjon, ved at det blir mulig å vekte valgene i analysene.
- Blokkdesign, som fanger opp pakkeeffekter (at verdsettingen av en pakke med forbedringer er forskjellig fra summen av de individuelle forbedringene), viser at faktorene relatert til reisen (sitteplass, reisetid og frekvens) verdsettes opptil 35 % lavere når de forbedres samtidig enn når de bare verdsettes enkeltvis.

Tabell S.1: Beregnede SC-verdsetninger (kroner pr time eller pr reise) av egenskapene fordelt på ulike design og datainnsamlingsmetoder

Kvalitetsfaktor	Design 1	Design 2	Design 3	Design 4	Design 5	Design 6	Hjemme-intervju	Egen-adm.	Alle
Reisetid, kr pr time	7	9	20	17	10	25	8	26	23
Ståplass, tillegg/time	15	17	17	15	11		16	12	13
Ventetid, kr pr time	77	88	69	61	48	44	82	66	68
Vektore, kr pr reise	2	1	-1	-1	0	2	1	0	0
Nødtelefon, kr pr reise	1	1	0	1	-3	-0	1	0	0
Overbygg, kr pr reise	1	0	-0	0	-5		1	-1	-1
Leskur, kr pr reise	2	2	1	1	1		2	1	1
Realtid, kr pr reise	1	0	1	1	3	2	1	2	1
Rutekart, kr pr reise	1	1	-1	0	-4	-1	1	-1	-1
Ant. respondenter	35	31	53	52	44	55	66	298	364

CV/TP-sekvensene

Vi har påvist følgende sammenhenger mellom egenskaper ved designet og CV/TP-verdsetninger av kvalitetsfaktorer:

- Respondenter har en tendens til å avrunde svarene sine. Dette påvirker resultatene og skaper skjevheter i verdsettingene.
- I spørsmål om kompensasjonskrav for en *forverring*, kan verdsettingen bli overdreven stor, fordi det ikke er spesifisert noen nedre grense for respondentens *betalingsvilje* for den forverrede reisen. Dette er en viktig årsak til at forverring tillegges større vekt enn forbedringer i CV/TP.
- Spørsmålsstillingen i CV/TP er direkte og har ingen elementer av avveining, noe som gjør metoden utsatt for protestvalg og strategiske svar. Dette er årsaken til flere observasjoner av 0 i *betalingsvilje* (gir høy verdsetting av en egenskap som forverres), en stor andel med 0 i *verdsetting*, og en del *overbud* som gir ekstremt høye verdsetninger av forbedringer.

Frisch-metoden

Mellom 14 og 18 % av respondentene valgte alternativet med lavest pris i alle FM-valgene, dvs. ville ikke godta høyere pris for å få hhv. redusert reisetid eller høyere frekvens. Vi har antatt at disse har 0 i betalingsvillighet. Sammenlignet med andelen respondenter med 0 i betalingsvilje i CV/TP, og som sorterer leksikografisk etter prisen i SC, er problemet med 0 i betalingsvillighet betraktelig redusert i FM.

Mellom 82 og 95 % av respondentene har en eller flere ganger valgt alternativet der de måtte avgi noe for å få en forbedring. Disse har altså veid endringer i godene mot hverandre. En stor andel respondenter foretar altså faktiske avveining i FM.

POG virker uegnet for verdsettingsstudier

Vi har ikke klart å gjøre nytte av materialet fra POG-sekvensene. POG er muligens egnet til å kartlegge hvilke kvalitetsfaktorer trafikantene prioriterer, men vi har ikke identifisert noen metode for å analysere POG-dataene som er tilfredsstillende for verdsettingsstudier.

Respondentene svarer delvis konsistent

Det er generelt godt samsvar mellom metodene når det gjelder hvilke egenskaper som er *viktigst* for de enkelte respondentene. Trafikantene har altså relativt klare prioriteringer som kommer til syne uavhengig av SP-metode.

Tabell S.2 viser et eksempel på verdsetninger som er beregnet i FM, CV/TP og samvalgene. Resultatene fra disse metodene er i samme størrelsesorden, men likevel med prosentvise store sprik.

Tabell S.2: Verdssetting av (reduisert) reisetid og (reduisert) tid mellom avgangene fra FM, CV/TP og SC. N=289-298¹⁾

	FM ¹⁾	CV/TP ¹⁾	SC
(Redusert) reisetid (kr/time)	32 – 39	19 – 20	23
(Redusert) tid mellom avgangene (kr/time)	32 – 34	22 – 28	34

¹⁾ Nedre grense er avveiningene når respondenter som tilkjenner ga betalingsvillighet på mer enn tre ganger taksten er ekskludert. I øvre grense er alle respondentene er inkludert.

Tester viser også at respondenter med spesielle preferanser svarer relativt konsistent på tvers av SP-metode. For eksempel har respondenter med høy betalingsvilje i CV/TP også tilkjenner en høyere verdssetting i samvalgene. Samtidig er det lite sammenheng mellom 0-verdssettinger i CV/TP og FM og leksikografiske svar i samvalgene. Vi har konkludert med at egenskaper ved spørsmålstillingen i de ulike metodene, og spesielt CV/TP har forårsaket dette.

Egenadministrasjon fungerer tilfredsstillende

Med bakgrunn i våre tester mener vi at egenadministrerte intervjuer er forsvarlig både faglig sett og ikke minst økonomisk. Vi anbefaler at det utarbeides mer brukervennlige løsninger for egenadministrerte intervjuer, f.eks. internett- eller windows-baserte løsninger, i den videre utviklingen av egenadministrerte intervjuer. Og vi vil igjen understreke nødvendigheten ved å supplere egenadministrerte intervjuer med hjemmeintervjuer for å sikre representasjon fra hele befolkningen.

Anbefalinger

- Leksikografi er et alvorlig problem i samvalgundersøkelser. Andelen leksikografi gir en god indikasjon på hvor godt samvalgdesignet har fungert, og bør alltid rapporteres.
- 0-svar og ekstreme verdssettinger er tilsvarende alvorlige problemer i CV/TP og FM. Disse andelene må derfor også rapporteres. FM har vist seg å ha lave andeler slike svar og er derfor funnet å være en bedre verdssettingsmetode enn CV/TP.
- Det er et vesentlig forbedringspotensial i å utvikle mer fleksibelt design av samvalgene, f.eks. ved å balansere nivåene i samvalgsekvensen ut fra betalingsvilje i andre typer SP-metoder. I denne studien er CV/TP brukt til dette. Dette gir en klar tendens til lavere andeler leksikografi. På grunn av metodiske svakheter ved CV/TP anbefales FM i stedet for CV/TP til å balansere samvalgsekvensen i fremtidige undersøkelser.
- Hjemmeintervjuer er svært ressurskrevende. Våre undersøkelser viser at egenadministrasjon, som er langt rimeligere, har fungert godt. Datainnsamlingsmetoden har hatt svært liten innvirkning på resultatene. Imidlertid må egenadministrerte intervjuer suppleres med hjemmeintervjuer for å sikre at alle grupper av befolkningen blir representert.
- Egenadministrerte intervjuer bør gjøres mer brukervennlige for å redusere frafallet. Vi anbefaler at det utvikles Windows- eller internettbaserte løsninger.
- Det anbefales økt bruk av kontrollspørsmål i alle typer SP-undersøkelser. Kontrollspørsmål kan avdekke årsaker til at respondentene tilkjenner ekstreme preferanser, og forekomsten av protestvalg og forenklete beslutningsrutiner.