



# Masteroppgave

Masterstudium i atferdsvitenskap – spesialisering i  
begreper og anvendelse

Juni 2023

Effekten av The Good Behavior Game på uønsket  
atferd hos elever i barneskolen: En metaanalyse

Kandidatnavn: Emma Askelien Moen

Emnekode: MALK5000

30 Studiepoeng

**Fakultet for helsevitenskap**

OSLO METROPOLITAN UNIVERSITY  
STORBYUNIVERSITETET

## **Forord**

Jeg vil med det første rekke en stor takk til min veileder, Børge Strømgren, for upåklagelig hjelp og veiledning det siste halvåret. Han har brukt mye av sin tid til praktisk hjelp, samtidig gitt god veiledning og tips til nye verktøy, metoder og litteratur. Hans faglige kompetanse og oversikt over området har gjort denne artikkelen gjennomførbar og lærerik. Jeg vil takke gode medstudenter, som gjennom pensumbaserte diskusjoner, og støtte har gjort de to siste årene minneverdige. Spesielt vil jeg takke medstudent og venninne, Vilde Ulfski, som har bidratt med kontroll av kritiske deler av denne prosessen. Jeg vil takke nysgjerrige kollegaer på jobb, som har stilt kritiske spørsmål, men vist interesse og optimisme for temaet. En ekstra stor takk til kollega, Zahra Ouchen, for korrekturlesing og tilbakemeldinger på artikkelen. Jeg vil til slutt takke gode venner, familie og nære som har motivert og støttet meg i prosessen. Utrolig takknemlig for alle rundt meg.

**Effekten av The Good Behavior Game på uønsket atferd hos elever i barneskolen:**

**En metaanalyse**

Emma Askelien Moen

Fakultet for helsevitenskap, institutt for atferdsvitenskap

OsloMet - Storbyuniversitetet

MALK5000

15. Juni 2023

### Sammendrag

Uønsket atferd i klasseromssituasjoner er et stadig problem i skolesystemet, som både er tidkrevende og forstyrrende for klasseledere og miljøet i klasserommet. Det er viktig for elever, og samfunnet at det finnes verktøy som er økonomiske og effektive, og som er lett tilgjengelige for klasseledere. The Good Behavior Game er et eksempel på et slikt tiltak, og som tidligere studier viser å ha god effekt på å redusere uønsket klasseromsatferd, og øke prososial atferd i klassemiljøet. Etter en systematisk tilnærming og søk i 5 databaser ble 8 studier med single-case design, som undersøkte effekten av GBG, innhentet. Artikkelen har undersøkt effekten GBG har på uønsket atferd gjennom en metaanalyse sammenfattet av effektstørrelser fra de ulike studiene. For å undersøke om GBG kan klassifiseres som en evidensbasert praksis, har det blitt foretatt en metodologisk kvalitetsvurdering på de innhentede studiene. Resultatene av metaanalysen viste høy samlet effektstørrelse, 2.93, på effekten av GBG på uønsket atferd. Dog var resultatene påvirket av høy grad av heterogenitet,  $I^2 = 92.9\%$ , og et stort konfidensintervall 95% [1.4, 4.45]. GBG ble klassifisert som en potensiell evidensbasert praksis. Av de 8 inkluderte studiene, oppfylte 2 kravene for å kunne bli vurdert for evidens, disse viste begge positiv effekt av GBG på uønsket atferd.

*Nøkkelord:* The Good Behavior Game, metaanalyse, single-case design, uønsket atferd, barneskole

### **Abstract**

Unwanted behavior in classroom situations is a constant problem in the school system. Which is both time-consuming and disruptive for classroom leaders and the environment in the classroom. It is important for students and society that there are tools that are economical and effective, and that are easily accessible and implemented by teachers. The Good Behavior Game is an example of such a measure, and as previous studies have shown to have a good effect on reducing unwanted classroom behavior and increasing prosocial behavior in the classroom environment. After a systematic approach and search in 5 databases, 8 studies with a single-case design which investigated the effect of GBG, were obtained. This article has investigated the effect GBG has on unwanted behavior through a meta-analysis summarized by effect sizes from these 8 studies. In order to investigate whether GBG can be classified as an evidence-based practice, a methodological quality assessment has been carried out on the studies obtained. The results of the meta-analysis showed a high overall effect size, 2.93, on the effect of GBG on unwanted behavior. However, the result was influenced by a high degree of heterogeneity,  $I^2= 92.9\%$ , and a large confidence interval 95% [1.4, 4.45]. GBG was classified as a potential evidence-based practice. Of the 8 included studies, 2 met all the requirements to be considered for evidence, both of which showed a positive effect of GBG.

*Keywords:* The Good Behavior Game, meta-analysis, single-case research, problem behavior, primary classroom

## **The Good Behavior Game for elever i barneskolen - en systematisk tilnærming**

The Good Behavior Game, ble introdusert som en klasseromsintervensjon, implementert av klasseleder, for å redusere forstyrrende atferd i klasserommet. The Good Behavior Game fungerer som et verktøy klasseledere kan bruke for å administrere betinget forsterkning til elevene sine, ofte i form av en belønning (Cooper, 2020, s. 720-721). Intervensjonens mål er at konsekvenser for gruppen har en effekt på individuelle kontingenter (Barrish et al. 1969; Flower et al. 2014). Denne artikkelen vil gjennom en systematisk tilnærming og metaanalyse utforske effekten The Good Behavior Game har på uønsket atferd i klasserommet. The Good Behavior Game vil videre i denne artikkelen bli forkortet til GBG.

Barrish, Saunders og Wolf (1969) utførte det som er identifisert som første studie der denne metoden ble brukt. Studien ble utført på en ordinær klasse på 4.trinn, som hadde flere barn som viste uønsket klasseromsatferd. Intervensjonen ble lagt opp som et spill i klasserommet: i den originale studien, ble klassen delt i to ulike lag, og samlet poeng ved å følge reglene læreren presenterte. Konsekvensen av fremvist uønsket atferd fra en elev i ett av lagene, førte til at hele laget mistet noen fordeler.

Studien Barrish et al. (1969) gjennomførte, undersøkte effekten av naturlige klasseromsforsterkere som kan brukes for å redusere uønsket klasseromsatferd og fremme prososial atferd i klasserommet. Økning i prososiale ferdigheter forbindes i GBG med gjennomføring av oppgaver, akseptere autoriteten i klasserommet, og forbedre konsentrasjonen til elevene (Barrish et al., 1969; Bowman-Perrott, 2016). Atkins et al., (2019, s. 25-26) forklarer at prososial atferd bidrar til å skape effektive, sosiale grupper, og utvikle individet innad i et fellesskap. Tingstrom et al. (2006, s. 225) argumenterer for at gruppeorienterte kontingenter fungerer like effektivt som individuelle kontingenter.

Det skal i forkant av implementeringen være avgjort hvilke atferder som går under betegnelsen uønsket atferd. Når klasseleder introduserer GBG som et spill for elevene, blir

spillet betinget med ulike regler. Reglene for spillet gjenspeiler i mange ulike studier, selv om det finnes noen variasjoner. Ofte brukte regler spesifiseres som følger (Bowman-Perrott et al., 2016):

(1) Ingen skal forlate plassen sin uten å ha fått tillatelse fra klasseleder. Tillatelse kan fås dersom eleven rekker opp hånden og venter på ordet. (2) Ingen skal sitte oppå pulten eller på side-personen sin plass. (3) Ingen skal flytte eller skyve på pulten sin. (4) Ingen skal forlate plassen sin for å prate med medelever, det gjelder også ved hvisking. (5) Dersom klasseleder sitter ved sin pult, kan elevene, en etter en, komme opp til klasseleder og stille spørsmål. (6) Ikke lov å snakke uten tillatelse, må rekke opp hånden først. (7) Ingen skal snakke mens hånden er hevet, elevene må vente på tillatelse først. (8) Ikke lov å hviske eller snakke med andre i klasserommet. (9) Ingen skal lage vokale lyder og bråk (Bowman-Perrott et al., 2016; Cooper et al., 2020, s. 720-721).

### **Modifikasjoner av The Good Behavior Game**

Tingstrom et al. (2006) har sammenfattet forskning gjort på GBG fra 1969 til 2002, og argumenterer for at den originale utformingen og ulike modifikasjoner av GBG er tilnærmet like effektive. Kriteriene som blir satt er opp til hvordan klasseleder, og eventuelt forskere, velger å organisere spillet i klasserommet. I denne artikkelen vil det være The Good Behavior Game som det undersøkes effekten av, dette inkluderer også ulike modifikasjoner av GBG i klasseromssituasjoner. Det beskrives i dag fire grunnelementer for GBG: Klasseromsregler, laginndeling, observasjon av regelbrytende- og regelfølgende atferd, og positiv forsterkning (AIR, 2021).

I den ordinære versjonen av GBG blir klassen delt inn i lag, og laget forsøker å oppnå poeng ved å følge reglene. Uønsket atferd blir definert i forkant av fasen, for eksempel som å snakke uten å få ordet, forlate plassen sin, forstyrre andre, etc. Laget som vinner får tildelt en type belønning (Hartman & Gresham, 2016; Lannie et al., 2007; Nolan et al., 2014; Peltier et

al., 2023). Klasseleder kan føre poeng på ulike måter gjennom spillet, for eksempel ved å bruke legoklosser eller andre telleverktøy. Lynne et al. (2017) tok i bruk et digitalt program som heter ClassDojo© som supplement til GBG. Klasseleder kan bruke dette for å tildele lagene poeng i programmet, og lagene kan følge med på tildelingen etter hver fase av spillet.

*Caught Being Good Game*, CBGG, beskrives som en positiv versjon av GBG. Denne versjonen bruker komponenter fra GBG, men tar i bruk en tidtaker for å tildele poeng. Poeng gis etter et variabel intervall-forsterkningskjema, og når alarmen går av skal klasseleder gi poeng til laget som på det tidspunktet fulgte reglene. Klasseleder kan i tillegg tildele poeng på slutten av hver fase (Wahl et al., 2016; Wright & McCurdy, 2012).

Strømgren og Sørheim (2015), utførte en studie som testet effekten av *The Good Behavior Board Game*, GBBG, som er en brettspill-variant av GBG utformet som en labyrint. Klasseleder spiller spillet i økter med klassen, i forkant er det definert hva regelbrudd er, og maksimalt antall regelbrudd klassen kan ha i løpet av økten. Ofte betinges spillet med maks 10 regelbrudd på 10 minutter. Dersom klassen oppfyller kravet, flytter klasselederen en brikke på brettet. Klassen mottar dermed en betinget forsterker, ved at klasselederen flytter brikken gjennom labyrinten. Når brikken når målet, får hele klassen tildelt en belønning, ofte trekkes det en lapp fra en "skattekasse" der belønningen står (Cipani, 2010).

## **Tidligere systematiske oversikter**

### ***Flower et al. (2014)***

Flower et al. (2014) undersøkte effekten av GBG på utfordrende atferd i deres oversikt av 22 studier gjort på effekter av GBG på utfordrende atferd i skolesetting. I oversikten ble det undersøkt virkningen GBG har på uønsket atferd, prosessen av intervensjonen og bruken av belønning på problematferd. Hovedfunnene i deres oversikt fant moderat til høy, umiddelbar effekt på uønsket atferd i klasseromssituasjoner ved bruk av måleenhetene Cohen's *d* for gruppedesign, og hierarkisk lineær modellering, HLM, for



single-case design. Målinger på basislinjen av forstyrrende atferd ble målt som høy, og etter implementering av GBG viste trenden i intervensjonsfasen en umiddelbar nedgang, og lite endring over tid. Samtidig påpekes det at implementeringen av tiltaket bør inneholde en belønningsprosedyre som utføres korrekt, og blir levert på riktig måte, for å være effektiv.

Oversikten undersøkte i tillegg ulike karakteristikk og grunner til differansene i resultatene fra gruppedesign og single-case studier. De beskrives som hvem som presenterte intervensjonen til elevene, behandlingsintegritet, ulike settinger, varighet av intervensjonen, og belønningene elevene fikk. GBG ble dermed påvist å være høyt effektivt på tvers av klasserom med mye utfordringer og uønsket atferd. Det påpekes at forstyrrende atferd i klasserommet, kan potensielt hindre en god arena for læring. Fordelen med klasseromsintervensjoner slik som GBG, er at det kan redusere tiden klasseleder må bruke for å håndtere forstyrrende atferd, og gi mer tid til læring (Flower et al., 2014).

### ***Bowman-Perrott et al. (2016)***

Bowman-Perrott et al. (2016) utførte en metaanalyse av single-case studier på elever fra 1.-12. trinn. Data som foreligger, er samlet inn fra 1580 deltakere, gjennom 21 studier. Analysen adresserte to forskningsspørsmål: (1) Den samlede effekten av GBG i økning av positiv atferd og reduisering av forstyrrende atferd. (2) Effekten av fem moderasjoner: EBD, risikostatus, forsterkningsfrekvens, interesseatferd, GBG format, og klassenivå. Måleenheten Bowman-Perrott et al. (2016) brukte for å regne effektstørrelsen var, Tau-U, deretter ble resultatet konvertert til Cohen's  $d$  for enklere å tolke og generere resultatet. De påpeker at det bør utvises forsiktighet ved konvertering mellom Cohen's  $d$  og Tau-U.

Resultatet som fremsto viste høy effektstørrelse av GBG på tvers av studiene (Cohen's  $d=1.99$ , 95% CI). Det var på forstyrrende og *off-task* atferd, det ble vist høyest effekt. Bowman-Perrott et al. (2016) beskriver at Tau-U for å måle effektstørrelse, baserer seg på målinger mellom A og B faser som ikke overlapper hverandre og at måleenheten kan

gi kontroll over konfunderende trender på basislinjen. Bowman-Perrott et al. (2016) brukte *What Works Clearinghouse*, WWC standarden, for å kvalitetsvurdere studiene, estimere publiseringsbias og vurdere effekt. Av metaanalysens 21 inkluderte studier, var det 5 studier som oppfylte kravene, og 9 oppfylte med forbehold, nok av kvalitetskravene til WWC standarden. Dette beskrives som et bidrag for å etablere GBG som en evidensbasert praksis (Kratochwill et al. 2010, i Bowman-Perrott et al., 2016).

### ***Smith et al. (2021)***

Smith et al. (2021) sin metaanalyse inkluderte 7 studier med til sammen 3700 deltakere, og gjennom en randomisert kontrollert studie, utforsket de effekten av GBG. Deltakerne var grunnskoleelever fra 5-11 år. Smith et al. (2021) fremstiller egne resultater, Hedges'  $g = .09-.32$ , som moderate i sammenligning med det Bowman-Perrott et al. (2016) kom frem til. Smith et al. (2021) påpeker at måleenheten Tau-U, er et relativt nytt verktøy for beregning av effektstørrelse, og det bør fremvises forsiktighet ved sammenligning og transformasjon fra Tau-U til Cohen's  $d$ .

Smith et al. (2021) konkluderer med at GBG sin effekt på uønsket klasseromsatferd, problemer mellom medstudenter og tilbaketrukkethet, overgikk sammenligningbetingelsene. Smith et al. (2021) påpeker forskjeller GBG har på kjønn, og at det bør vurderes om en klasse er passende for å motta GBG med tanke på klassen som helhet, enkelt-elevenes behov og atferd. Smith et al. (2021) fremviser at GBG er effektiv innenfor spesifikke utfall, som for eksempel atferdsproblematikk og leseforståelse, fremfor uoppmerksomhet.

### **Evidensbasert praksis**

Anvendt atferdsanalyse har et overordnet mål, at endringene som utføres er av sosial viktighet. Det er ønskelig å forstå og forberede menneskelig atferd, gjennom et felles fokus, mål og metoder. En grunnleggende målsetting innenfor feltet av anvendt atferdsanalyse er å ha empirisk støtte for de tiltakene som gjennomføres. Dersom et tiltak eller intervensjon kan

karakteriseres som evidensbasert praksis, må den ha blitt vist å være effektiv av et betydelig antall fagfellevurderte studier (Cooper et al., 2020, s. 18, 96).

Cook et al. (2015) påpeker at et tiltak eller en intervensjon ikke kan karakteriseres som evidensbasert på bakgrunn av en enkelt studie med positiv effekt, og det er utarbeidet en veileder for en systematisk tilnærming for kvalitetsvurdering og vurdering av evidens, *Standards for Classifying the Evidence Base of Practices in Special Education*. Veilederen ble utarbeidet, og senere revidert av *Council of Exceptional Children*, CEC, og blir beskrevet i Cook et al. (2015) og Royer et al. (2017). Cook et al. (2015) beskriver i tillegg hvordan den systematisk har blitt gjennomgått av 23 fagpersoner med et snitt på over 14 års erfaring innenfor feltet av spesialopplæring i skolesystemet, og testet IOE gjennom en pilotstudie.

CEC standarden for å vurdere evidens består av to områder: (a) hvordan utforske metodologisk forsvarlighet gjennom ulike kvalitetsindikatorer, og (b) kriterier for å identifisere evidensbasert praksis for utvalgte studier innenfor gruppedesign og single-case design (Cook et al., 2015). Vurdering om hva som er evidensbasert praksis er nødvendig for å kunne utforske hvilke tilgjengelige tiltak som er mest effektive, og at forholdet mellom forskning og praksis samstemmer. Innenfor atferdsanalyse foretrekkes det design der det blir foretatt gjentatte målinger av samme atferd, hos de samme deltakerne. Dette for å avdekke en funksjonell relasjon mellom tiltak og endring i atferd (Cooper et al., 2020, s. 20, 76).

Mye av tidligere forskning på GBG har brukt en variant av single-case forskningsdesign, som betraktes som et pålitelig forskningsdesign (Hedges et al., 2013). For å se på effekten av GBG blir det ofte gjort målinger på basislinjen i en klassetime, uten at tiltaket er introdusert. På denne måten oppnås det en stabil basislinje for hvor mange ganger uønsket atferd oppstår i klasserommet. Data fra samme gruppe eller klasse blir evaluert, og en overvekt av studiene som allerede er utført viser reduserende endring på forekomsten av interesseatferden fra basislinje til tiltaksfase (Tabell 1 i Tingstrom et al., 2006, s. 227-237).

## Grunnlaget for metaanalysen

Tidlig innsats står i fokus for barn som står i fare for å bli hengende etter, og det er gunstig å iverksette tiltak for dette allerede de første årene på skolen. Allerede tidlig i grunnskolen begynner enkelte barn å falle utenfor skolesystemet. Konsekvensen kan være at utfordringene får utvikle seg over tid, og kan føre til at unge står i fare for å ikke fullføre grunnskole eller videregående opplæring. Barn lærer fortere dersom intervensjonen forekommer tidlig, og vil ofte ha større nytte av tiltakene (Udir, u.å). Studier av GBG som har blitt utført viser høy kortsiktig-og langsiktig effekt på uønsket atferd, og en økning i prososial- og samarbeidsatferd hos deltakerne (Gamble & Strain, 1979, i Tingstrom et al., 2006, s. 245).

Tidligere litteratur viser at GBG er effektiv også på verbal og fysisk vold, og andre variasjoner av utfordrende atferd som kan oppstå i skolesituasjoner, atferden som påpekes her er mangel på overholdelse av regler, motstridende ytringer og hyperaktivert atferd (Bowman-Perrott et al., 2016). Bowman-Perrott et al. (2016) og Smith et al. (2021) har gjennomført metaanalyser som undersøkte effekten av GBG, og beskriver noen ulikheter i resultatene. Forfatter mener derfor det er interessant å utføre en metaanalyse, og produsere et resultat som kan supplere det som har blitt utforsket tidligere.

GBG som intervensjon beskrives som et enkelt og allsidig verktøy, som ikke krever store ressurser for å gjennomføre. Tidligere empiri tyder på at konsekvenser gitt på gruppekontingenser gir flere fordeler, blant annet krever det mindre tid og fokus fra klasseleder på å identifisere og levere forsterkere. Samtidig som det viser positiv påvirkning på interaksjoner mellom elevene og samarbeidet i klasserommet (Gresham & Gresham, 1982). Publisert faglitteratur gir en antydning av at intervensjonen er lite studert og utbredt i det norske skolesystemet. Forfatter ønsker derfor at denne artikkelen kan bidra til et større fokus på GBG som et tiltak i skolen, dersom effekten viser å være like høy som først antatt.

Formålet med systematiske tilnærminger, og metaanalyser, er å kunne rapportere en samlet effektstørrelse for en intervensjon, gjennom å sammenfatte flere studier gjort på et område (Borenstein, et al. 2017; Liberati et al., 2009). Med denne metaanalysen er det ønskelig å foreta en systematisk gjennomgang av studier, med formålet om å undersøke effekten GBG har på uønsket atferd hos barn på barneskoler. Denne artikkelen vil undersøke dette ved å, (a) identifisere kortsiktig reduserende effekt av GBG, for å kunne forebygge uønsket og forstyrrende atferd i klasserommet, (b) vurdere om GBG kan klassifiseres som en evidensbasert praksis for å redusere uønsket atferd i klasseromssituasjoner, (c) sammenligne effekten av GBG fra allerede gjennomførte studier ved å foreta en metaanalyse av effektstørrelsene.

### **Metode**

I forkant av søkeprosessen ble det gjort bakgrunnsøk på litteratur om GBG, både enkeltstudier, metaanalyser og andre litteraturbidrag. Dette for å opparbeide et empirisk grunnlag innenfor temaet. Det ble brukt to tidligere metaanalyser for å få en oversikt over fremgangsmåten for en lignende artikkel som denne. Smith et al. (2021) og Bowman-Perrott et al. (2016) ble i hovedsak brukt til dette. PRISMA sjekklister for systematiske oversikter og metaanalyser ble brukt for disposisjon og innhold (Liberati et al., 2009), samt APA-manualen for formelle krav til en slik artikkel (American Psychological Association, 2020).

### **Databaser og søkestrategi**

Et systematisk søk i databasene ble gjennomført for å ha en systematisk tilnærming til dataene. Søkene ble delt inn i tre elementer med utgangspunkt i forskningsspørsmålet, som var utarbeidet tidlig i skriveprosessen. De tre hovedelementene for søket ble som følger: *The good behavior game*, *disruptive behavior* og *classroom*. Det ble gjort søk i totalt fem ulike databaser: OVID for tilgang i databasen PsycInfo. EBSCOhost som leverandør for søk i ERIC, Education Source og Academic search ultimate. Det siste søket ble gjennomført i Web

of Science. Grunnlaget for å velge ut disse databasene var for å innhente resultater på tvers av ulike fagfelt innenfor atferdsvitenskap, psykologi og pedagogikk, og Web of Science ble brukt med grunn i at den gir tilgang til forskningsdatabaser som består av flere disipliner.

Emneord og nøkkelord ble notert underveis, og de samme ordene ble brukt på tvers av databasene. Dersom det dukket opp nye søkeord ble disse testet i foregående database, for å se om det kunne inkluderes der og. I noen tilfeller skjedde dette, da ble søket redigert for å inkludere søkeordene. Dette var under utvidelsen av søkeord om atferdsproblematikk, da det var mange variasjoner innenfor temaet. Booleske operatører, *AND* og *OR*, ble brukt i søkene for å kombinere, begrense eller utvide søkeordene. I alle databasene var det mulig å begrense søket med fagfelleverderte artikler og språk. Treff ble overført til forfatter sin EndNote.

### **Begrepsavklaring**

*Baseline.* Oversatt til basislinjen eller basislinje i denne artikkelen, forklares som målinger gjort av atferd før tiltak igangsettes (NAOB, 2023).

*Target behavior.* En beskrivelse av elevene sin atferd som ikke er ønskelig i klasserommet, og atferdskategorier vi ønsket å endre. Oversatt til interesseatferd (Cooper, 2020, s. 69). Det finnes ulike begreper som henviser til slik atferd i studiene, for eksempel, forstyrrende-, problem-, aggressiv atferd, og hyperaktivitet, regelbrudd, *out-of-seat-behavior*. Alle atferder som i klasseromssituasjoner kategoriseres som variasjoner av disse, vil videre i artikkelen bli benevnt som uønsket atferd.

*Klasseleder.* Artikkelen beskriver individet som har ledelsen i klasserommet og den som iverksetter intervensjonen under eksperimentet, som klasseleder, dersom ikke annet er spesifisert. Rollen til klasseleder er i denne artikkelen kontaktlærer, men kan i noen tilfeller være forskere eller andre ansatte på skolen. Et overordnet begrep, brukt av Cook et al. (2015) og Royer et al. (2017) er intervensjonsagent.

*Gruppekontingenser.* Litow & Pumroy (1975) identifisert tre former for gruppekontingenser: uavhengig, avhengig, og gjensidig avhengig. I klasseromssituasjoner er det ønskelig å gi forsterkning, for eksempel i form av belønning til gruppen som helhet, ikke til enkeltindivider. Avhengig gruppekontingenser innebærer at gruppen fremdeles kan motta forsterker, så lenge noen individer innenfor grupper har møtt kriteriet satt for fremvist atferd. Gjensidig avhengig gruppekontingenser krever at hele gruppen møter kriteriene for at alle skal motta forsterkeren. Artikkelen tar utgangspunkt i avhengig eller gjensidig avhengig.

### **Inklusjons- og eksklusjonskriterier**

Tabell 1, fremstiller en oppsummert sjekklister for alle inklusjonskriteriene. Den ble brukt som veileder da forfatter og medstudent skulle gjennomgå og kontrollere søket. Forfatteren satt krav til at studiene var publisert i en *peer-reviewed journal*, heretter fagfellevurdert, innenfor felt som psykologi, pedagogikk, spesialpedagogikk vitenskap og atferdsvitenskap. Andre metaanalyser, studentarbeid, litteraturoversikter eller sammendrag ble ekskludert.

Forfatter satte krav for publiseringsdato til maksimalt 20 år tilbake, per dags dato tilsvarende det, 2003. Dette av hensyn til endringer i skolesystemet og perspektivet på barnets status i samfunnet. Denne endringen oppsto som en konsekvens av implementeringen av FNs Barnekonvensjon (1989) og tiden etter det. Treff ble ekskludert dersom de var publisert på andre språk enn engelsk, svensk, dansk eller norsk.

Studien refererer til GBG eller ulike versjoner av GBG som bruker GBG sine grunnelementer etter AIR (2021), som uavhengig variabel, med grunn i at litteratur tilsier at ulike modifikasjoner kan ha lignende effekt (Tingstrom et al., 2006). Forfatter ønsket at intervensjonen skulle være betinget av konsekvenser gitt til hele klassen eller en klasse delt inn i flere lag eller grupper.

Deltakerne som gjennomgikk intervensjonen ble begrenset til grunnskolebarn, omtrentlig alder 6-12 år, i ordinære skoleklasser. Dermed ble andre aldersgrupper ekskludert, samt studier gjort i spesialklasser, spesialskoler og barnehager. Ordinære klasser med enkeltbarn som viste atferdsproblematikk ble ikke ekskludert. I ordinære klasserom er det ofte læreren ansvar for klasseromsledelse, derfor ble det satt krav til at klassens lærer, eller tilsvarende voksen, implementerte intervensjonen i klasserommet. Forfatter satt som kriterium at observasjon og innhentet data ble registrert av enten, klasseledere, forskere eller studenter med opplæring i det.

Atferden, avhengig variabel, ble beskrevet som forstyrrende eller problematferd som er hemmende for klassemiljøets læring, trygghet og sosiale forhold. Det ble satt krav til at atferden hadde blitt observert og gjort målinger av. Dataene som forekom måtte kunne ekstraheres, visuelt analyseres, og være mulige å kalkulere effektstørrelsen av til en senere metaanalyse. Forfatter begrenset forskningsområdet til å vurdere effekten GBG har på uønsket atferd i klasserommet. Målinger av akademiske prestasjoner, prososial atferd eller intervensjoner som har forsket på andre problemområder slik som fedme eller røyking, ble ekskludert.

Av hensyn til individuelle forskjeller hos elevene og at ordinære klasser ofte har ulike sammensetninger av elever, ønsket forfatteren at målinger fra basislinje til tiltaksfase var utført hos samme individer. Dermed satt forfatter krav til at analysen kun skulle inkludere studier med et forskningsdesign som hadde gjort målinger av basislinje og i tiltaksfase på samme subjekter. Forskningsdesign innenfor dette kriteriet kunne for eksempel være single-case, i form av reverseringsdesign, AB- design, *alternate treatment design*, herav ATD, eller *multiple baseline design*, heretter MBD. På grunn av begrensning i forskningsdesign ble studier utført med kontrollgrupper ekskludert (Cooper, 2020, s. 199, 223; Hedges et al., 2013).



## **Seleksjonsprosessen**

Forfatter brukte EndNote sin duplikat-funksjon, for å gjennomføre fjerning av duplikater. Treffene ble videre sortert etter forfattere i EndNote, deretter utførte forfatter en manuell gjennomgang for å kontrollere duplikater. Fjerning ble gjort dersom forfatter, årstall, og tittel stemte overens. Denne delen ble utført uten medstudent. Ved usikkerhet rundt kriteriene for ekskludering, ble treffet stående og gjennomgått i en senere fase.

Forfatter gjennomførte en manuell screening av titler og sammendrag for å ekskludere deltakere som ikke gikk under ordinære grunnskoleklasser. Ordinære grunnskoleklasser er ofte 1.-7.trinn eller alderen 6-12 år. Det ble under avansert søk i EndNote brukt søkeordene: *high school, college, prison, kindergarten, adults, sport, autism* og *special education*. Dette var gjennomgående ord i overskrifter der ekskluderingskriteriene var raskt identifiserbare for fjerning. Treff ble fjernet dersom tittel ikke stemte overens med inkluderingskriterier.

Etter screening av overskrifter, ble sammendragene til resterende treff lest for å identifisere inkluderingskriteriene. Andre settinger som for eksempel treningssituasjoner, virtuelle spill, alkoholdriking, spesialskoler og spesialundervisning ble fjernet. Andre typer intervensjoner og forskningsområder ble ekskludert etter gjennomgang av sammendragene. Dersom inkluderings- eller ekskluderingskriterier ikke ble beskrevet i sammendraget eller titlene, ble de ikke fjernet. Deretter ble treffene som gjensto lest metoden til, for å kunne identifisere hvilke artikler som kvalifiserte seg med tanke på design og type interesseatferd. Det siste steget var å lese studiene i fulltekst for å identifisere inklusjonskriteriene, og studiene som kunne ekstrahere sammenlignbare data til en senere metaanalyse.

## **Kontrollprosedyre**

Ved hjelp av en medstudent fra samme studieretning ble seleksjonsprosessen kontrollert. Treffene som gjenstod etter fjerning av duplikater lå i en samlet mappe i

EndNote. Kontrollen ble gjennomført randomisert, der artiklene ble sortert etter forfattere, og medstudenten leste tittel og sammendraget til hver fjerde artikkel i oversikten. Dette for å oppnå *interrater agreement*, eller inter-observatør-enighet, heretter IOE, mellom score på inkluderingskriteriene (Christensen et al., 2015, s. 156). Medstudenten, som hadde fått inkluderingskriterier, lagde en oversikt over hvilke som oppfylte eller ikke oppfylte disse kriteriene. Denne prosessen gjorde forfatter og medstudent uavhengig av hverandre. Etterpå gikk forfatter og medstudent gjennom oversikten, for å sammenligne resultatene. Ved uenighet i resultatene kunne veileder drøftes med, for å komme til en enighet.

## **Kvalitetsvurdering**

### ***Deskriptiv vurdering***

Karakteristikker ved variabler i studiene ble beskrevet og vurdert. Dette for å senere kunne identifisere variabler i studiene som kunne vært påvirkningsfaktorer ved resultatene. Disse variablene ble identifisert: (a) alder/klassestrinn, (b) antall deltakere, (c) landet studien ble utført i, (d) variant av GBG som ble implementert, (e) interesseatferden som ble beskrevet i studien, (f) type eksperimentelt design brukt, (g) klasseleder sin rolle.

### ***CEC's kvalitetsindikator, CEC QI***

Metodologisk kvalitetsvurdering ble gjort etter en standardisert beskrivelse av kriteriene etter CEC QI, i Cook et al. (2015) og Royer et al. (2017). Appendix A viser en oppsummert oversikt over kvalitetsindikatorerne, og beskrivelse av kriteriene som ble vurdert. Forfatteren og medstudenten scoret studiene etter de åtte kvalitetsindikatorerne som ble beskrevet, scorene ble ført inn uavhengig av hverandre i Excel matrisen (Lane et al., 2014). Samme medstudent, som kontrollerte søkeprosessen, ble brukt i denne prosessen. Excel matrisen beregnet verdiene som ble lagt inn separat, og kalkulerte IOE.

Kun kriterier og data som var relevante for artikkelen ble innhentet og gitt score på, totalt på single-case studier var dette 22 komponenter. Forfatter og medstudent la inn om

studien innvilget (1) eller ikke innvilget (0) kriteriene til hver komponent. Ved uenighet diskuterte forfatter og medstudent resultatene, og gjennomgikk punktet på nytt til avsluttende koding samstemte. For å innvilge (1) kriteriet måtte det ved kontroll mellom forfatter og medstudent, bli fremvist utdrag av tekst fra studien (Lane et al., 2014; Royer et al., 2017).

### ***CEC's vurdering av evidensbasert praksis, CEC EBP***

Basert på verdiene som ble lagt inn i QI-matrisen og kodet (Lane et al., 2014), ble verdiene for sammenlagt-, og absolutt score regnet for alle studiene. Dersom alle QI komponentene blir møtt, tilsvarer dette en absolutt score på 8.0. En modifisert tilnærming vil være å evaluere studiene som oppfyller 80% vektet-kodingskriterium. Verdier som scoret over minimumskravet, på 6.4 poeng eller høyere, kan vurderes for evidens. Det påpekes at dersom studien møter minimumskravet av 80% vektet-kodingskriterium eller absolutt score, men scoret 0 på komponent 6.5 i QI, vil det ikke være etablert en funksjonell relasjon mellom avhengig- og uavhengig variabel. Studien kan dermed ikke bli evaluert som evidensbasert og vil automatisk bli fjernet fra en videre vurdering av evidens (Royer et al., 2017).

Cook et al. (2015) beskrev fem ulike klassifiseringer for å vurdere evidens. Disse beskrivelsene ble brukt for å vurdere om GBG etter denne analysen, kunne klassifiseres som en evidensbasert praksis. For at en studie kunne vurderes for evidens måtte studien ha oppfylt absolutt- eller sammenlagt score etter CEC QI, samt scoret (1) på komponent 6.5. Dermed blir studiene vurdert med tilstrekkelig metodologisk kvalitet for å gjennomføre evidensvurdering. For en mer detaljert beskrivelse kan Cook et al. (2015) anvendes. Kriteriene som videre beskrives gjelder for analyser av single-case studier.

(1) Evidensbasert praksis, krever minimum fem studier med tilstrekkelig metodologisk kvalitet med positiv effekt, og minimum 20 deltakere totalt fra studiene. Forholdet mellom studiene som viser positiv effekt og studiene som viser nøytral eller blandet effekt må være 3:1. Ingen av studiene kan vise negative effekter (Cook et al., 2015).

(2) Potensielt evidensbasert, må ha støtte fra 2-4 studier med tilstrekkelig metodologisk kvalitet som viser positiv effekt, og forholdet mellom studiene som viser positiv effekt og studiene som viser nøytral eller blandet effekt må være 2:1. Ingen av studiene kan vise negative effekter (Cook et al., 2015).

(3) Blandet evidens, må enten kriteriene for evidensbasert eller potensielt evidensbasert praksis være oppfylt, men forholdet mellom studiene med tilstrekkelig metodologisk kvalitet som viser positiv effekt og studier som viser nøytrale eller blandete effekter er under 2:1. Det kan forekomme negative effekter i noen av studiene, men antallet må være lavere enn studiene med positive effekter (Cook et al., 2015).

(4) Utilstrekkelig evidens, ingen av de øvrige klassifiseringene er oppfylt. Det finnes ikke tilstrekkelig evidens hos inkluderte studier, mer forskning kreves (Cook et al., 2015).

(5) Fremvisning av potensielle negative effekter. Antall studier som viser negative effekter er høyere enn antallet studier som fremviser positive resultater av studiene med tilstrekkelig metodologisk kvalitet (Cook et al., 2015).

### **Visuell analyse**

Cook et al. (2015) argumenterer for at å kunne klassifisere potensielle effekter i single-case studier må forfatteren utføre visuell analyse for å konkludere om det er etablert en funksjonell relasjon, mellom intervensjonen og resultatene for hver case. For å identifisere en funksjonell relasjon må grafene utforsket for endring i (a) level, gjennomsnittlig score innenfor samme fase (b) trend, retningen til linjen innenfor fasen (c) variasjon av score rundt trendlinjen, (d) umiddelbarheten av effekten i ny fase, og (e) om datapunktene på tvers av faser overlapper. Forfatter ønsket å se en lavere forekomst av uønsket atferd i tiltaksfasen på samtlige studier. I tillegg utførte medstudent en visuell analyse av de 8 inkluderte studiene, for å sikre IOE av klassifisering av effekt.

Cook et al. (2015) påpeker at de som sammenfatter studiene kun bør visuelt analysere og inkludere data som artikkelen skal omhandle. Etter å ha gjennomført en visuell analyse av relevante caser i studiene, identifiserte forfatter og medstudent om det var en funksjonell relasjon mellom avhengig og uavhengig variabel. Studiene sine grafer ble undersøkt i de ulike fasene, og klassifisert med positiv (+), negativ (-), eller nøytral/blandet (N/B) effekt.

En positiv effekt fremvises ved at det er vist en positiv funksjonell relasjon mellom tiltaket og avhengig variabel, og at endringen er sosial valid for minimum 75% av inkluderte caser. Det kan ikke etableres en positiv effekt dersom noen av casene viser skadelig effekt. Negative effekter fremkommer dersom det vises en funksjonell relasjon mellom tiltak og avhengig variabel, og skadelig effekt ved minimum 75% inkluderte caser. Skadelige effekter eksemplifiseres ved en økning i den atferden som det ønskes en reduksjon av, eller en reduksjon i atferd studien ønsker mer av (Cook et al., 2015).

For at en studie skal klassifiseres for å ha nøytrale eller blandede effekter, skal verken kriteriet for positive eller negative effekter bli møtt. Med utgangspunkt i design, kvantitet, og klassifisering av effekt, kan forfatter kunne konkludere om studiene kan beskrives som evidensbasert, potensielt evidensbasert, har blandede effekter, eller om studien har utilstrekkelig med evidens, eller negative effekter (Cook et al., 2015).

### **Metaanalyse: Ekstraksjon av data og effektstørrelser**

Forfatteren fikk hjelp til det praktiske av veileder til å ekstrahere og generere effektstørrelser fra de ulike studiene. Alle studiene hadde et single-case design der data ble fremstilt i grafer. Datapunkter fra grafene som fremviste resultater for uønsket atferd, ble innhentet av forfatteren og lagt inn i en Microsoft Excel- fil. Appen SingleCasES (Pustejovsky et al., 2023), ble brukt for å analysere hver fase isolert og samlet. Appen estimerte effektstørrelse, som er sammenlignbar med Cohen's  $d$  og Hedges'  $g$  i gruppedesign (Hedges et al., 2013).

Konvertering fra Cohen's  $d$  til Hedges'  $g$  ble gjort i appen basert på antall datapunkter i serien. Ved senere metaanalyse bør en robust metode brukes for estimering av samlet effektstørrelse, derfor ble en randomisert effekt modell med estimatoren *Restricted Maximum Likelihood* med trunkert Knapp-Hartung justering (Borenstein et al., 2021, s. 243-249) brukt i SPSS28. Prediksjonsintervallet ved få antall studier, kan være så bredt at det blir lite informativt, og vil krever større justering (Borenstein et al., 2017). Forrest plot ble beskrevet av Borenstein et al. (2009, s. 366) som et hjelpemiddel for å studere samlet effektstørrelse og konfidensintervallet i kontekst. Dette ble brukt for å enklere se variasjoner i statistikken.

Heterogenitet ble kalkulert etter  $I^2$ , verdien kan gi et anslag av variasjon i den estimerte effekten mellom studiene.  $I^2$  kvantifiserer spredning i effektstørrelsen blant de inkluderte studiene, og høyere prosentvis verdi representerer høyere nivåer av heterogenitet. Eksempelvis kan nivåer av heterogenitet tolkes etter observerte verdier av  $I^2$ :  $I^2=0\%$  tilsvarer ingen observerte nivåer av heterogenitet,  $I^2=25\%$ , lavt nivå,  $I^2=50\%$ , moderate nivåer, og  $I^2=75\%$  høye nivåer. Høyere nivåer av heterogenitet kan medføre vanskeligheter med å sammenligne effektstørrelse (Borenstein et al., 2017). Ved få inkluderte studier anbefales det ikke å beregne eventuell publikasjonsskjevhet (Borenstein et al., 2009, s. 284).

## Resultater

### Utvelgelse av studier

Appendiks B, Tabell 1B-3B, gir en oversikt over søketreff og resultater fra søk i leverandørene OVID, EBSCOhost, og Web of Science. Dato for gjennomføring av søk var 17. og 18. januar 2023. Etter gjennomført søk i fem databaser ble det eksportert totalt 1272 treff til EndNote. Forfatteren inkluderte 1 studie (Strømgren og Sørheim, 2015), som ble identifisert utenfor databasesøk. Figur 1, viser en visuell fremstilling av seleksjonsprosessen gjennom et flytdiagram (Liberati et al., 2009). Fjerning av duplikater ble gjort med duplikat-funksjonen i EndNote ( $n=431$ ). Deretter foretok forfatter en manuell gjennomgang for å

kontrollere duplikater ( $n=96$ ). Fjerning ble gjort dersom forfatter, årstall, og overskrift stemte overens. Av artiklene som ble screenet ( $n=756$ ), og gjennomgått i fulltekst ( $n=42$ ), oppfylte 8 studier inkluderingskriterier og ble tatt med til kvalitetsvurdering og senere metaanalyse. IOE var 100% mellom forfatter og medstudent etter at medstudent kontrollerte 25% av treffene.

### **Kvalitetsvurdering av inkluderte studier**

#### ***Deskriptiv vurdering***

Tabell 2 viser til en oversikt over de ulike studiene sine karakteristikk som var relevante etter inkluderings- og ekskluderingskriteriene forfatter bestemte før metaanalysen. Totalt fra syv av studiene, inkluderte det 256 deltakere, med ett avvik der Strømgren og Sørheim (2015) ikke hadde oppgitt antall deltakere. Nolan et al. (2014) oppga totalt antall elever som deltok. Resterende studier oppga antall elever per klasse. Samtlige studier presenterte hvilket klassetrinn deltakerne befant seg i. Nolan et al. (2014) beskrev alder og klasse, dermed ble “*Beginners*” inkludert da dette var barn på 6 år, som var innenfor inkluderingskriterium for alder og ordinære klasseromssituasjoner. Alle studiene ble publisert på engelsk og utført i ulike stater i USA, med unntak av Strømgren og Sørheim (2015), som ble publisert på norsk og utført i en norsk barneskole.

Videre viser Tabell 2 forskningsdesign for alle studiene, og ulike variasjoner. Det ble foretatt målinger av basislinje, A, og minimum én intervensjonsfase, B i alle caser. Hartman og Gresham (2016), Peltier et al. (2023), Wahl et al. (2016), og Wright og McCurdy (2012), hadde i tillegg C-faser. Data fra C-fasene ble ikke innhentet og en del av metaanalysen. Kun A og B betingelser ble inkludert, der B-fasens betingelse var GBG. Alle studiene hadde gjort målinger i ordinære klasserom, der klasseleder var den som implementerte og utførte intervensjon. Unntaksvis studien til Peltier et al. (2023) der forskeren, elevene og lærer hadde noen økter hver i intervensjonsfasen, kun data fra øktene til forsker og lærer ble innhentet til senere metaanalyse.

### ***CEC og visuell analyse***

Tabell 3 viser score etter visuell inspeksjon av studiene. Her forekommer det at 4 av studiene kunne fremvise positiv effekt, 4 av studiene viste N/B effekt. Ingen av studiene hadde negative effekter. IOE mellom forfatter og medstudent etter visuell analyse og klassifisering av effekt var 100%. Videre presenteres IOE-score, kalkulert til 98.3% gjennomsnittlig etter komponenter. Gjennomsnitt etter studier lå på 98.66% (variasjonsbredde fra 96.4-100%). Uenighetene lå ved komponent 6.7 for Peltier et al. (2023), og Wright og McCurdy (2012), og komponent 7.4 for Wahl et al. (2016). Det ble utregnet score på indikatorene, 1.0-8.0. Ved verdi (1), ble samtlige av komponentene i indikatoren innvilget. Her viste avslutningsvis koding at ingen av studiene nådde absolutt score på 8.0 poeng. 5 av 8 studier møtte kravet for over 80% vektet-kodingskriterium, med variasjonsbredde fra 6.67 til 7.67 (Hartman & Gresham, 2016; Lannie & McCurdy, 2007; Lynne et al., 2017; Wahl et al., 2016; Wright & McCurdy, 2012).

Figur 2 viser en oversikt over scoring av komponentene i CEC QI, og hvilke komponenter hver enkelt studie innvilget (1) eller ikke innvilget (0) basert på total score. 22 av 28 indikatorer ble scoret på for hver enkelt studie: av 176 komponenter ble 137 scoret som innvilget (1). Dette tilsvarte 77.8% av komponentene (Lane et al., 2014). Hartman og Gresham (2016), Wahl et al. (2016), og Wright og McCurdy (2012) scoret 0 på komponent 6.5 fordi de ikke kunne fremvise tre målinger av effekt på data som skulle innhentes i analysen: herunder ble én til to caser ekskludert på grunn av alder eller klassesertrin som gikk utenfor inklusjonskriteriet på 6-12 år eller barneskoleklasser. Strømgren & Sørheim (2015) scoret (0) på komponent 6.5 da de ikke oppfylte 75% kriteriet for positiv effekt av tiltaket: 1 av 3 caser viste N/B effekt, tilsvarende 66.7% caser med positiv effekt. Den nøytrale effekten ble begrunnet med manglende behandlingstrokap. 4 av 8 studier scoret (1) på 6.5.

### **Effektstørrelse og metaanalyse**



Tabell 4 oppgir verdier for effektstørrelser, standardfeil, og 95 % CI. Samlet effektstørrelse var 2.93,  $SE = 0.645$ ,  $p = 0.003$ , 95 % konfidensintervall [1.4, 4.45], 95 % prediksjonsintervall [-1.27, 7.12]. Utregning for sannsynlig heterogenitet i datasettet,  $Q=56.49$ ,  $p=.001$ ,  $\tau^2 = 2.258$ ,  $I^2 (\%) = 92.9$ . Antall datapunkter ( $n=$ ) med varians fra 10 (Wright & McCurdy, 2012), 13 (Hartman & Gresham, 2016), 17 (Lannie & McCurdy, 2007), 25 (Wahl et al., 2016) til 60 (Peltier et al., 2023; Strømgren & Sørheim, 2015), 64 (Lynne et al., 2017), 66 (Nolan et al., 2014). Figur 3 viser *forest plot*, der hver studies effekt og konfidensintervall samt vekt i den samlede effektstørrelse vises.

### Diskusjon

Forfatter ville med denne artikkelen undersøke effekten GBG har på uønsket atferd i klasseromssituasjoner. Tidligere analyser har demonstrerer høy og umiddelbar reduserende effekt. Analysens formål var å replikere tilsvarende resultater gjennom kvalitetsvurderinger av studier som har utforsket effekten av GBG, og gjennom en metaanalyse. Samlet effektstørrelse etter metaanalyse viser 2.93, som kan vurderes som høy effektstørrelse. Derimot vurderes resultatene som usikre, på grunn av høy heterogenitet mellom studiene. Resultatene fra gjennomført metodologisk kvalitetsvurdering, og en senere evidensvurdering, vurderes GBG som en potensiell evidensbasert praksis, med støtte fra to studier som fremviste positive effekter.

### Søkeresultater

Valg av inklusjon- og eksklusjonskriterier kan ha medført at potensielle og relevante studier har blitt utelatt fra denne artikkelen. I Cooper et al. (2020, s. 96) trekkes det frem at det må fremvises positiv effekt fra et betydelig antall fagfelleverderte studier, for at en intervensjon kan karakteriseres som en evidensbasert praksis. Forfatter stilte krav til fagfelleverderte studier, og gjennomførte søk i fem databaser for tilgang gjennom OsloMet. Studier med potensielt relevante resultater som kunne påvirket analysen, kan ha blitt oversett.

*Kindergarten* kan i noen stater eller land, også være 1. klassering, og elever fra 5-6 år, men som direkte blir oversatt til barnehage i det norske språket. Dermed ble data fra *kindergarten* utelatt, for å ikke inkludere data som er hentet fra barnehager, som ofte har en annerledes struktur enn i ordinære klasserom. Nolan et al. (2014) er et unntak det de hadde beskrevet strukturen og alder hos deltakerne i *beginners*-gruppen, og forfatter vurderte casen som innenfor inkluderingskriteriene. Forfatteren satte ikke krav til lengden av intervensjonen. Ved stor variasjon i lengde, for eksempel om intervensjon i løpet av noen få uker eller strekker seg over et skoleår, påvirker dette mengden elevene har blitt utsatt for intervensjonene på.

I motsetning til tidligere metaanalyser satt forfatter publiseringskrav. Et kriterium for GBG i dag, er positiv forsterkning som en grunnpilar i spillet (AIR, 2022; Cooper et al., s. 720-721), sammenlignet med Barrish et al. (1969) som trakk fra poeng dersom det var individer som brøt reglene. Grunnen til at publiseringsår ble et kriterium, for å sikre et likere syn på opplæring i skolen som i dag. Dersom studier og analyser gjennomført mellom 1969-2003 har brukt andre metoder enn det som er vanlig i dag, kan elevene være utsatt for en annen pedagogikk og forsterkningsprosedyre, enn hva som ble brukt før 2000-tallet. Dette kan medføre at elever i dag respondere ulikt på intervensjonen enn hva de gjorde før.

### **Deskriptiv vurdering**

Det vises noe variasjon mellom de gjennomførte studiene. Syv av de inkluderte studiene brukte kun den ordinære varianten av GBG, eller denne i kombinasjon med andre. Alle studiene ble utført hos de nedre trinnene i grunnskolen, og ble implementert, i hovedsak av, klassens lærer. En forskjell som trekkes frem er antall caser de ulike studiene gjennomførte, der to studier viste data fra én case, og de resterende gjennomførte to eller tre caser. Dette kan påvirke tyngden til studiene, og kan ha påvirket den metodologiske kvaliteten. Samtidig forekommer det noe variasjon i design, der én studie brukte ATD, to

brukte MBD, og fem brukte reverseringsdesign. Samtlige studier gjorde målinger av basislinjen, og hadde minimum én A-fase og én B-fase. Det er en mulighet for at designene med C kriterium kan ha innvirkning på resultatet. Av hensyn til at C-fasene ikke er inkludert i metaanalysen, kan resultatene allikevel være påvirket av eventuelle rekkefølgeeffekter.

### **Vurdering av kvalitet og evidens**

Forfatter er kjent med at det finnes andre standarder for å vurdere kvaliteten til studier, for eksempel WWC (Bowman-Perrott et al., 2016), men CEC ble valgt med bakgrunn i at det er en anerkjent standard, utformet for spesialtiltak i undervisning. Resultatene av metodologiske vurderinger kan tilsi moderat til høy metodologisk kvalitet på studiene der 77.8% av komponentene i CEC QI ble innvilget, og fem av studiene lå over 80% vektet-kodingskriterium, tilsvarende en sammenlagt score over 6.4.

Totalt hadde 2 av 8 studier tilstrekkelig metodologisk kvalitet og scoret (1) på komponent 6.5, herunder Lannie og McCurdy (2007), og Lynne et al. (2017). Begge studiene kunne fremvise positiv effekt, og kriteriet for potensiell evidensbasert praksis oppfylles. Artikkelen kunne ikke avdekke evidensbasert praksis, da dette hadde krevd fremvist positiv effekt og tilstrekkelig metodologisk kvalitet hos 5 av de inkluderte studiene. Samtidig viste ingen av de 8 inkluderte studiene å ha negativ effekt, som kan gi en indikasjon på at det er nødvendig med ytterligere forskning for at det skal kunne konkludere med en evidensbasert praksis basert på denne analysen (Cook et al., 2015; Royer et al., 2017).

Videre scores det lavt på komponent 6.3, og dette kan begrunnes i at designet valgt for metaanalysen, er ugunstig dersom det forekommer 2 eller flere A-faser, noe som 5 av de inkluderte studiene bruker. Deltakerne som blir utsatt for flere faser, vil i den andre A-fasen ha tilgjengelighet til intervensjonen. Dette er viktig for å replikere effekten av tiltaket, men medfører en svakhet for vurdering av metodologisk kvalitet. Det samme gjelder for komponent 2.2 der 4 av 8 studier scorer (0), som kan begrunnes med at deltakerne er

ordinære skoleklasser, der det ikke er sikkert noen elever har en påvist diagnose eller annen risiko, og for mange av disse studiene kan det dermed være ment unødvendig å inkludere rapportering av risikostatus.

### **Kontrollprosedyre og utvelgelse**

IOE mellom forfatter og medstudent ble testet ved flere kritiske steder gjennom analysen. Først for å kontrollere selekteringsprosessen, dette ble utført med 100% IOE. Videre ble det testet IOE i kvalitetsvurderingen av studiene, og ved visuell analyse av effekt før vurdering av evidens. IOE viste å være relativt høy i alle fasene, med variasjonsbredde fra 96.4-100% ved score på CEC QI, og 100% etter visuell analyse og CEC EBP.

### **Vurdering av samlet effektstørrelse og metaanalyse**

Antall datapunkter fra studiene varierte fra 10 til 66 og dette viser en relativt stor variasjon i omfanget av studiene som er inkludert i artikkelen. Ved metaanalyser av små studier, kan det være vanskelig å fremstille nøyaktige resultater da enkelte variasjoner kan gi betydelig forskyvning i samlet effektstørrelse (Borenstein et al., 2017). Større studier har større vekt, dermed bidrar mest i den samlede effektstørrelsen og påvirkning av resultater.

Samlet effektstørrelse var 2.93, med en signifikant  $p$ -verdi på 0.003, og et bredt 95% konfidensintervall mellom 1.4 og 4.45. Dette antyder at det er en stor grad av heterogenitet i datasettet, som også er indikert til en høy  $I^2$ -verdi på 92.9% og en signifikant  $p$ -verdi på 0.001 for  $Q$ -testen. Heterogenitet representerer faktiske forskjeller i studiene, som ikke kan tilskrives av tilfeldigheter (Borenstein et al., 2017). Ingen av studiene krysset 0 på CI 95% nedre grense, som kan tilsi en signifikant relasjon (0.26-2.47). 95 % konfidensintervallet [1.4, 4.45] indikerer at vi kan være 95 % sikre på at den sanne populasjonsparametre ligger mellom disse verdiene. På samme måte indikerer 95 % prediksjonsintervallet [-1.27, 7.12] at vi kan være 95 % sikre på at en fremtidig observasjon vil falle innenfor dette området.

Det er viktig å merke seg at det ikke ble beregnet eventuell publikasjonsskjevhet på grunn av det begrensede antallet studier som ble inkludert i analysen (Borenstein et al., 2009, s. 284). Det betyr at det kan være risiko for at noen studier ikke er inkludert i datasettet, for eksempel studier med negative resultater som kan ha blitt utelatt eller vanskeligere å finne. Dette er noe som må tas i betraktning i tolkning av resultatene som presenteres i analysen.

### **Tidligere oversikter**

Samtlige av tidligere metaanalyser viser at GBG har større effekt på å redusere *off-task* og forstyrrende atferd, enn å øke oppmerksomheten hos elevene. Sammenlignet med tidligere oversikter kan ikke denne artikkelen konkludere med like sterke bevis, som det Bowman-Perrott et al. (2016) demonstrerte, med at GBG viser høy effektstørrelse på uønsket atferd og etablering av evidensbasert praksis. Det kan diskuteres om ulike verktøy brukt for beregning av evidens og kvalitetsvurdering kan være grunnen til differanse i Bowman-Perrott et al. (2016) sine resultater og denne analysen.

Artikkelen kan dog diskutere likere resultater som det Smith et al. (2021) og Flower et al. (2014) konkluderte med, at GBG ikke virker å være like effektivt som først identifisert, basert på resultater fra analysene. Flower et al. (2016) konkluderer med at GBG har moderat til høy effekt på uønsket atferd, men fremviser noen variasjoner mellom single-case design og studiene med gruppedesign. Det bør tas hensyn til at variasjonene i resultatene hos tidligere analyser, kan være påvirket av konverteringen mellom ulike måleenheter, da dette bør utføres med forsiktighet (Bowman-Perrott et al., 2016; Smith et al., 2021)

### **Videre forskning**

GBG blir innledningsvis beskrevet som en evidensbasert intervensjon som er designet for å forbedre uønsket klasseromsatferd. Resultatene av metaanalysen viser imidlertid en høy grad av heterogenitet blant de inkluderte studiene som har undersøkt effektene av GBG.

Videre forskning bør derfor undersøke hva som ligger bak denne heterogeniteten, for

eksempel ulike kontekstuelle faktorer eller varierende behandlingsintegritet. Det kan anbefales å ha fokus på den metodologiske scoringen etter CEC, derav hvilke komponenter som senker kvaliteten på CEC QI og EBP. For eksempel på indikator 6.0 for indre validitet, og 5.0 for behandlingsintegritet, scorer flere av de inkluderte studiene (0). Videre forskning kan forbedre kvaliteten i rapportene ved økt fokus på disse områdene (Se Figur 2). Flower et al. (2014) argumenterer at implementeringen av tiltaket bør inneholde en belønningsprosedyre som utføres korrekt for å være effektiv. Dette samstemmer med blant annet Strømgren og Sørheim (2015) som ikke oppfylte kravet for positiv effekt, på grunn av N/A hos én case, og forklarte dette med mangel på behandlingsintegritet.

Forfatter ønsket å inkludere studier utført i det norske skolesystemet, etter søk i Norsk Tidsskrift for Atferdsanalyse, ble én studie funnet. Med en overvekt av amerikanske studier kan diskuteres om det amerikanske skolesystemet er overførbart til det norske eller nordiske systemet. Mer forskning og studier av GBG i Norden kunne vært interessant (Strømgren & Sørheim, 2015). Mange studier ble ekskludert på grunn av andre forskningsområder. En annen mulig retning for fremtidig forskning kan være å undersøke langtidseffekter av GBG. De fleste studiene i metaanalysen hadde korte oppfølgingsperioder, og det er derfor uklart om forbedringene i atferd vil vedvare over tid.

### **Sosiale og etiske hensyn**

Gjennomføring av denne artikkelen har fulgt anbefalte retningslinjer for systematiske tilnærminger, PRISMA (Liberati et al. (2009), samt vurdering av kvaliteten til innhentede studier er gjort etter CEC (Royer et al., 2017). APA-manualen (2020) er brukt for å sikre kildehenvisning og kildekritikk. Intervensjonen som er studert i denne artikkelen, er omtalt som en mindre inngripende, enn mange andre tiltak, som er et prinsipp i atferdsanalytisk praksis. Det er derfor et ønske for forfatteren at det utforskes mer på dette området, og om

dette er en metode som er forsvarlig å igangsette, samtidig som at den oppfyller krav for en potensiell evidensbasert praksis (Cooper et al., 2020, s. 18, 96).

I denne metaanalysen har vi undersøkt effekten av GBG på uønsket atferd i klasserommet. I tillegg til å evaluere effektiviteten av intervensjonen, har det også vurdert intervensjonens sosiale validitet ved å score på CEC QI, der alle inkluderte studier scorer (1) på indikatoren for sosial validitet. Et viktig aspekt ved sosial validitet er intervensjonens evne til å forbedre klasseromsatferd på en måte som er akseptabel og bærekraftig for både elever og klasseledere. GBG er et lite inngripende tiltak som legger vekt på positiv forsterkning og samarbeid mellom elever, og kan være godt egnet for implementering i klasserommiljøer dersom det implementeres riktig og tilrettelegges klassens behov (Atkins et al., 2019, s. 25-26; Cook et al., 2015; Cooper et al., 2020, s. 18; Flower et al., 2014).

### **Konklusjon**

Resultatene viser høy samlet effektstørrelse, men bør tolkes med forsiktighet. Det kan på bakgrunn av effektstørrelse, 2.93, vise at GBG virker å ha reduserende effekt på uønsket atferd. Dog blir det vanskelig å konkludere med en samlet effektstørrelse etter denne analysen, da det fremkommer en  $I^2$ -verdi på 92.9%. Med det vurderes betydelig sannsynlighet for heterogenitet i datasettet som bør tas i betraktning ved tolkning av resultatene. I lys av disse resultatene vil det være nødvendig å undersøke de metodologiske forskjellene mellom studiene som er inkludert i analysen, for å identifisere eventuelle faktorer som kan påvirke heterogeniteten i datasettet. Etter metodologisk kvalitetsvurdering, og vurdering av evidens, ble GBG klassifisert som potensiell evidensbasert praksis. Det ble ikke fremvist forholdsmessige studier med positiv effekt og tilstrekkelig metodologisk kvalitet for å klassifisere GBG som en evidensbasert praksis.

### Referanser

- American Institute for Research “AIR” (2009-2021). *About the good behavior game*. Hentet 25.januar 2023 fra [https://goodbehaviorgame.air.org/about\\_gbg.html](https://goodbehaviorgame.air.org/about_gbg.html)
- American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7.utg.). <https://doi.org/10.1037/0000165-000>
- Atkins, P. W. B., Wilson, D. S. & Hayes, S. C. (2019). *Prosocial: Using evolutionary science to build productive, equitable, and collaborative groups*. New Harbinger Publications.
- Barnekonvensjonen. (1989). *FNs konvensjon om barnets rettigheter* (20-11-1989). Lovdata. [https://lovdata.no/dokument/NL/lov/1999-05-21-30/KAPITTEL\\_8#KAPITTEL\\_8](https://lovdata.no/dokument/NL/lov/1999-05-21-30/KAPITTEL_8#KAPITTEL_8)
- Barrish, H. H., Saunders, M., Wolf, M. M. (1969). Good Behavior Game: effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis*, 2, 119–124. <https://doi.org/10.1901/jaba.1969.2-119>
- Borenstein, M, Higgins, J. P. T., Hedges, L. V. & Rothenstein, H. R. (2017). Basics of meta-analaysis:  $I^2$  is not an absolute measure of heterogeneity. *Res. Syn. Meth.*, 8(5), 5-18. <https://doi.org/10.1002/jrsm.1230>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2021). *Knapp-Hartung Adjustment*. Wiley Online Library. <https://doi.org/10.1002/9781119558378.ch26>
- Borenstein, M., Hedges, L.V., Higgins, J.P.T., & Rothstein, H.R. (2009). *Introduction to meta-analysis*. Wiley.
- Bowman-Perrott, L., Burke, M. D., Zaini, S., Zhang, N., & Vannest, K. (2016). Promoting Positive Behavior Using the Good Behavior Game: A Meta-Analysis of Single-Case Research. *Journal of Positive Behavior Interventions*, 18(3), 180–190. <https://doi.org/10.1177/1098300715592355>



Cipani, E. (2010). *The class-wide Good Behavior Board Game*. Hentet fra

<https://eric.ed.gov/?id=ED512078>

Christensen, L. B., Johnson, R. B. & Turner, L. A. (2015). *Research methods, design, and analysis* (12. utg.). Pearson Education Limited.

Cook, B. G., Buysse, V., Klingner, J., Landrum, T. J., McWilliam, R. A., Tankersley, M., & Test, D. W. (2015). CEC's Standards for Classifying the Evidence Base of Practices in Special Education. *Remedial and Special Education*, 36(4), 220–

234. <https://doi.org/10.1177/0741932514557271>

Det Norske Akademisk Ordbok, NAOB. (2023). *Basislinje*. Det norske akademi for språk og litteratur. <https://naob.no/ordbok/basislinje>

Flower, A., McKenna, J. W., Bunuan, R. L., Muething, C. S., & Vega, R. (2014). Effects of the Good Behavior Game on Challenging Behaviors in School Settings. *Review of Educational Research*, 84(4), 546–571. <https://doi.org/10.3102/0034654314536781>

Gresham, F. M., & Gresham, G. N. (1982). Interdependent, dependent, and independent group contingencies for controlling disruptive behavior. *The Journal of Special Education*, 16, 101–110. <https://doi.org/10.1177/002246698201600110>

Hartman, K. & Gresham, F. (2016) Differential Effectiveness of Interdependent and Dependent Group Contingencies in Reducing Disruptive Classroom Behavior. *Journal of Applied School Psychology*, 32, 1-23.

<https://doi.org/10.1080/15377903.2015.1056922>

Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2013). A standardized mean difference effect size for multiple basislinjen designs across individuals. *Research Synthesis Methods*, 4, 324–341. <https://doi.org/10.1002/jrsm.1086>

- Lane, K. L., Common, E. A., Royer, D. J., & Muller, K. (2014). *Group comparison and single-case research design quality indicator matrix using Council for Exceptional Children 2014 standards* [Upublisert verktøy]. Ci3T. <https://www.ci3t.org/practice>
- Lannie, A. L., & McCurdy, B. L. (2007). Preventing Disruptive Behavior in the Urban Classroom: Effects of the Good Behavior Game on Student and Teacher Behavior. *Education and Treatment of Children*, 30, 85–98. <https://doi.org/10.1353/etc.2007.0002>
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., Clarke, M., Devereaux, P. J., Kleijnen, J. & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med*, 6(7), Artikkel e1000100. <https://doi.org/10.1371/journal.pmed.1000100>
- Litow, L., & Pumroy, D. K. (1975). A brief review of classroom group-oriented contingencies. *Journal of Applied Behavior Analysis*, 8(3), 341–347. <https://doi.org/10.1901/jaba.1975.8-341>
- Lynne, S., Radley, K.C., Dart, E.H., Tingstrom, D.H., Barry, C.T., Lum, J.D.K., (2017). Use of a technology-enhanced version of the good behavior game in an elementary school setting. *Psychology in the Schools*, 54, 1049–1063. <https://doi.org/10.1002/pits.22043>
- Nolan, J. D., Filter, K. J., & Houlihan, D. (2014). Preliminary report: An application of the Good Behavior Game in the developing nation of Belize. *School Psychology International*, 35(4), 421–428. <https://doi.org/10.1177/0143034313498958>
- Peltier, W., Newell, K.L., Linton, E., Holmes, S.C. and Donaldson, J.M. (2023), Effects of and preference for student- and teacher-implemented good behavior game in early elementary classes. *Journal of Applied Behavior Analysis*, 56(1) 216-230. <https://doi.org/10.1002/jaba.957>

- Pustejovsky, J. E., Chen, M., Grekov, P. & Swan, D. M. (2023). *Single-case effect size calculator* (Versjon 0.7.1) [nettapplikasjon]. <https://jepusto.shinyapps.io/SCD-effect-sizes/>
- Royer, D. J., Lane, K. L., & Common, E. A. (2017). *Group comparison and single-case research design quality indicator matrix using Council for Exceptional Children 2014 standards: Standards overview and walk-through guide* [Upublisert verktøy]. Ci3T. <https://www.ci3t.org/practice>
- Smith, S., Barajas, K., Ellis, B., Moore, C., McCauley, S. & Reichow, B. (2021). A Meta-Analytic Review of Randomized Controlled Trials of the Good Behavior Game. *Behavior Modification*, 45(4), 641–666.  
<https://doi.org/10.1177/0145445519878670>
- Strømgren, B., & Sørheim, D. G. (2015). Evaluering av the Good Behavior Board Game, en variant av the Good Behavior Game. *Norsk Tidsskrift for atferdsanalyse*, 42(1), 1-19.  
<http://www.nta.atferd.no/journalissue.aspx?IdDocument=539>
- Tingstrom, D. H., Sterling-Turner, H. E., & Wilczynski, S. M. (2006). The Good Behavior Game: 1969-2002. *Behavior Modification*, 30(2), 225-253. <https://doi.org/10.1177/0145445503261165>
- Wahl, E., Hawkins, R. O., Haydon, T., Marsicano, R., & Morrison, J. Q. (2016). Comparing Versions of the Good Behavior Game: Can a Positive Spin Enhance Effectiveness? *Behavior Modification*, 40(4), 493–517. <https://doi.org/10.1177/0145445516644220>
- Wright, R. A., & McCurdy, B. L. (2012). Class-Wide Positive Behavior Support and Group Contingencies: Examining a Positive Variation of the Good Behavior Game. *Journal of Positive Behavior Interventions*, 14(3), 173–180. <https://doi.org/10.1177/1098300711421008>

**Tabell 1***Sjekkliste for inkludering av studier til videre analyse*

Inklusjonskriterium	Beskrivelse
<i>Peer-reviewed</i>	Fagfellevurdert artikkel.
Publiseringsår	Maksimalt 20 år tilbake, publisert mellom 2003-2023.
Språk	Engelsk, Norsk, Svensk eller Dansk.
Uavhengig variabel	Intervensjon: GBG, variasjoner av GBG, GBG i kombinasjon med andre verktøy. Grunnelementer i AIR (2021) og gruppekontingenser.
Deltakere	Grunnskolebarn i ordinære klasser, omtrentlig alder 6-12 år.
Implementering	En lærer, forsker eller tilsvarende voksen implementere intervensjonen.
Observatør	Dataene er hentet fra lærer eller ekstern observatør med opplæring.
Avhengig variabel	Atferden beskrives som variasjoner av uønsket i klasseromssituasjoner og er hemmende for klasse miljøet.
Effekt	Effektstørrelsen i studien må kunne ekstraheres og brukes i metaanalysen.
Forskningsdesign	Designet er utformet slik at det har blitt foretatt måling av basislinje og tiltaksfase hos samme individer. Variasjoner innenfor single-case design eller <i>Multiple baseline design</i> .

*Merknad.* Oversikt over inklusjonskriterier for metaanalyse.

**Tabell 2***Deskriptive karakteristikk ved inkluderte studier*

Studie	Deltakere	Antall	Land	Intervensjon	Interesse- atferd	Design	Klasse- leder
Hartman & Gresham (2016)	3.klasseA	22	USA	GBG	Forstyrrende	ATD	Lærer
	3.klasseB	16	USA	GBG	Forstyrrende	ATD	Lærer
Lannie & McCurdy (2007)	1.klasse	22	USA	GBG	Forstyrrende	ABAB	Lærer
Lynne et al. (2017)	4.klasseA	27	USA	GBG m/ClassDojo	Forstyrrende	ABAB	Lærer
	1.klasse	19	USA	GBG m/ClassDojo	Forstyrrende	ABAB	Lærer
	4.klasseB	19	USA	GBG m/ClassDojo	Forstyrrende	ABAB	Lærer
Nolan et al. (2014)	Beginners			GBG	<i>Target</i>	MBD ABAB	Lærer
	2.klasse	Totalt 32	USA	GBG	<i>Target</i>	MBD ABAB	Lærer
	4.-5.klasse			GBG	<i>Target</i>	MBD ABAB	Lærer
Peltier et al. (2023)	1.klasse	24	USA	GBG: lærer implementert	Forstyrrende	ABABC	Lærer / forsker
	2.klasse	19	USA	GBG: lærer implementert	Forstyrrende	ABABC	Lærer / forsker
Strømgren & Sørheim (2015)	4.trinn	I.O	NOR	GBBG	Regelbrudd	MBD AB	Lærer
	7.trinn	I.O	NOR	GBBG	Regelbrudd	MBD AB	Lærer
	1.trinn	I.O	NOR	GBBG	Regelbrudd	MBD AB	Lærer
Wahl et al. (2016)	1.-2.trinn	16	USA	GBG/CBGG	Forstyrrende	ACBCB	Lærer
	2.trinn	20	USA	GBG/CBGG	Forstyrrende	ACBCB	Lærer
Wright & McCurdy (2012)	4.klasse	20	USA	GBG/CBGG	Forstyrrende	ACAB	Lærer

*Merknad.* I.O.=ikke oppgitt data; ATD=Alternating Treatment Design; A=Basislinje;

B=Tiltaksfase, GBG; C=Tiltaksfase med annen betingelse enn GBG; MBD=Multiple

baseline design; Beginners=6. åringer. Target behavior=interesseatferd, i denne studien

beskrives som å snakke uten å få ordet, å sitte urolig og å tulle.

**Tabell 3***Total score av CEC QI og EBP*

Deltakere (n)	38	22	65	32	43	I.O.	36	20
1.0 Kontekst	1	1	1	1	1	1	1	1
2.0 Deltakere	1	.5	1	1	.5	1	.5	1
3.0 Intervensjonsagent	1	1	1	.5	1	1	1	1
4.0 Beskrivelse av tiltaket	.5	1	1	.5	1	1	1	1
5.0 Prosedyreintegritet	1	1	1	.67	.67	.33	.67	1
6.0 Indre validitet	.83	.5	.67	.5	.5	.5	.5	.5
7.0 Avhengig variabel	.8	.8	1	.8	1	1	1	1
8.0 Dataanalyser	1	1	1	1	1	1	1	1
IOE (%)	100 %	100%	100 %	100 %	96.4 %	100 %	96.4 %	96.4 %
Absolutt score	5.0	5.0	7.0	3.0	4.0	5.0	5.0	6.0
80% vektet-kodingskriterium	7.13	6.80	7.67	5.97	6.17	6.33	6.67	7.00
Prosent QI oppfylt	89.1%	85.0%	95.9%	74.6%	77.1 %	79,1 %	83.4%	87.5%
Effekt	N/B	+	+	+	+	N/B	N/B	N/B
	Hartman & Gresham (2016)	Lannie & McCurdy (2007)	Lynne et al. (2017)	Nolan et al. (2014)	Peltier et al. (2023)	Strømngren & Sørheim (2015)	Wahl et al. (2016)	Wright & McCurdy (2012)

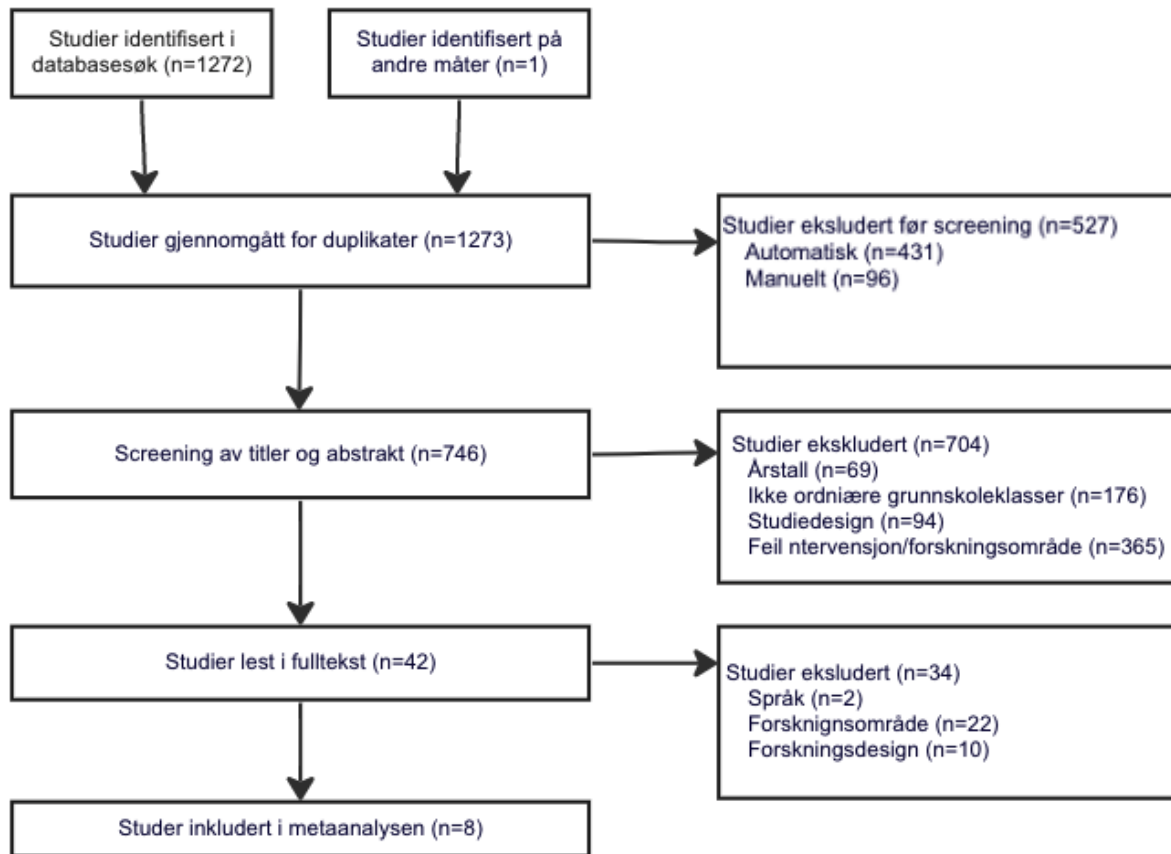
*Merknad.* IOE= inter-observatør enighet mellom forfatter og medhjelper; absolutt score viser antallet indikatorer der alle komponenter scoret til oppfylt verdi; 80% vektet-kodingskriterium= summen av andelen komponenter som scoret som oppfylt pr. indikator; QI= kvalitetsindikator, her oppgitt i prosent av 8, basert på sammenlagt score; I.O=ikke oppgitt; effekt etter gjennomført visuell analyse; N/B=Nøytral eller blandet effekt; +=positiv effekt.

**Tabell 4***Effektstørrelser for hver studie*

Studie	N	ES	Est.	SE	CI_nedre	CI_øvre
Hartmann & Gresham (2016)	13	SMD	6.16	1.28	3.65	8.67
Lannie & McCurdy (2007)	17	SMD	1.47	0.55	0.40	2.54
Lynne et al. (2017)	64	SMD	2.14	0.31	1.53	2.75
Nolan et al. (2014)	66	SMD	3.37	0.40	2.58	4.16
Peltier et al. (2023)	60	SMD	1.95	0.31	1.34	2.57
Strømgren & Sørheim (2015)	60	SMD	0.83	0.29	0.26	1.39
Wahl et al. (2016)	25	SMD	4.38	0.73	2.95	5.81
Wright & McCurdy (2012)	10	SMD	5.46	1.38	2.76	8.17
Samlet effektstørrelse	315	SMD	2.93	0.64	1.4	4.45

*Merknad.* N=Antall datapunkter; ES= Mål for effektstørrelse; SMD=*Standardized*

*Mean Difference*; Est.= Verdier av ES; SE=Standardfeil; CI=grense for konfidensintervallet (95%).

**Figur 1***Flytdiagram seleksjonsprosessen*

*Merknad.* Fremstillingen gir en oversikt over årsaker til eksklusjon. Antall studier

identifisert ved litteratursøk i databasene, og i Norsk Tidsskrift for Atferdsanalyse.

Antall studier ekskludert og inkludert i metaanalysen. Inspirert av PRISMA flytskjema

(Liberati et al., 2009).



**Figur 2**

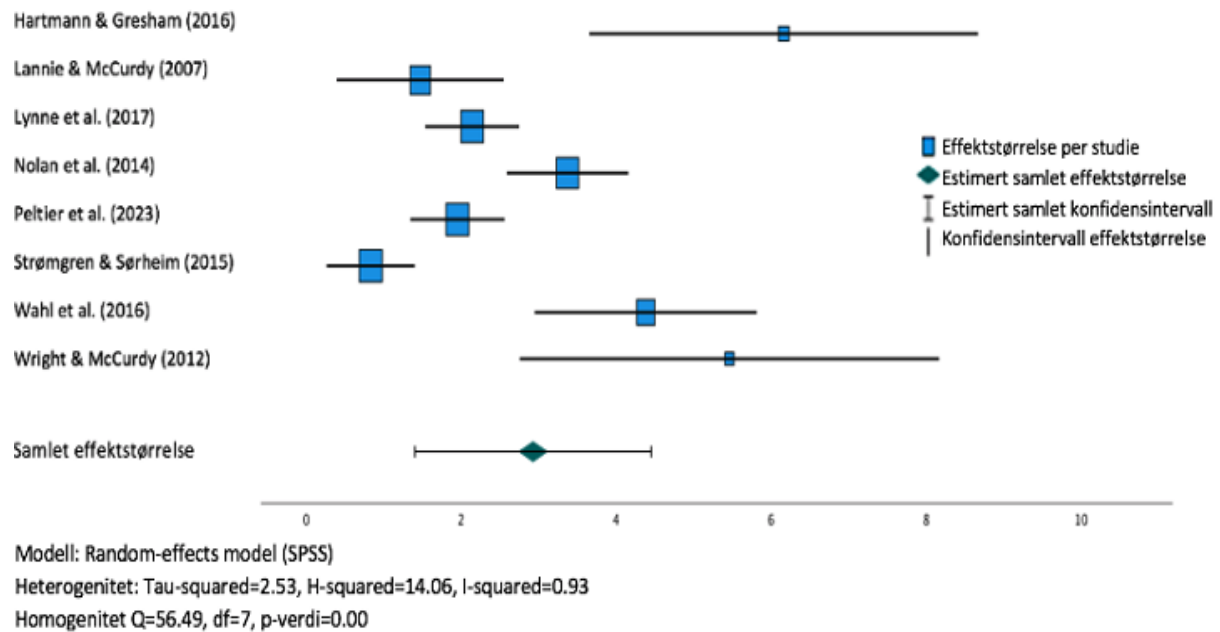
Resultater på CEC's kvalitetsindikator for de inkluderte studiene

8.3	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
8.2	1	1	1	1	1	1	1	1
8.1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
7.6	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
7.5	1	1	1	1	1	1	1	1
7.4	0	0	1	1	1	1	1	1
7.3	1	1	1	1	1	1	1	1
7.2	1	1	1	0	1	1	1	1
7.1	1	1	1	1	1	1	1	1
6.9	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
6.8	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
6.7	1	1	1	0	0	0	1	0
6.6	1	0	1	1	1	1	1	1
6.5	0	1	1	1	1	0	0	0
6.4	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
6.3	1	0	0	0	0	1	0	0
6.2	1	0	0	1	0	1	1	1
6.1	1	1	1	0	1	0	0	1
5.3	1	1	1	0	0	0	0	1
5.2	1	1	1	1	1	1	1	1
5.1	1	1	1	1	1	0	1	1
4.2	0	1	1	0	1	1	1	1
4.1	1	1	1	1	1	1	1	1
3.2	1	0	1	1	1	1	1	1
3.1	1	1	1	1	0	1	0	1
2.2	1	1	1	0	0	0	1	0
2.1	1	1	1	1	1	1	1	1
1.1	1	1	1	1	1	1	1	1
	Hartman & Gresham (2016)	Lannie & McCurdy (2007)	Lynne et al. (2017)	Nolan et al. (2014)	Peltier et al. (2023)	Strømgren & Sørheim (2015)	Wahl et al. (2016)	Wright & McCurdy (2012)

*Merknad.* N/A= Ikke anvendbar. 1=Innvilget. 0=Ikke innvilget. 1.1 Beskrivelse av setting; 2.1 Beskrivelse av deltakere; 2.2 Beskrivelse av risiko; 3.1 Rollen til klasseleder; 3.2 Oppløring; 4.1 Detaljer i prosedyren; 4.2 Materialer brukt; 5.1 Behandlingsintegritet; 5.2 Rapportering av pålitelige målinger; 5.3 Målinger i samtlige faser; 6.1 Kontroll og manipulasjon; 6.2 Rapportering av basislinjen; 6.3 Utilgjengelighet til intervensjonen; 6.5 Eksperimentell effekt ved tre demonstrasjoner; 6.6 Tre datapunkter i basislinjen; 6.7 Trusler mot indre validitet; 7.1 Sosial validitet; 7.2 Avhengig variabel; 7.3 Effekt for alle resultater; 7.4 Minst tre målinger i alle faser; 7.5 IOE målinger; 8.2 Graf som presenterer resultater som kan visuelt analyseres.

**Figur 3**

*Forrest plot, effektstørrelse for hver studie og samlet effektstørrelse*



*Merknad.* Forrest plot med effektstørrelser for hver studie, og samlet effektstørrelse, fra data som er innenfor inklusjonskriteriene. Linjene viser konfidensintervall for hver studie, og nederst samlet konfidensintervall. Størrelse på firkant samsvarer med vekten hver studie har på samlet effektstørrelse. Modellen er generert i SPSS28.

## **Appendiks A**

### **Council for Exceptional Childrens kvalitetsindikatorer og kriterier for score**

Kvalitetsindikatorer og kriteriene for score som ble vurdert i artikkelen (1) eller (0), er gjengitt etter beskrivelser fra Cook et al. (2015). Komponenter som gjelder for gruppedesign, ble i artikkelen ikke vurdert og blir derfor ikke beskrevet. Dersom det ikke blir nevnt under komponenten, gjelder beskrivelsen både studier for single-case design og gruppedesign.

### **Kontekst og setting**

#### **1.1. Beskrivelse av egenskaper i omgivelsene**

Studien beskriver nødvendig informasjon om kritiske egenskaper av kontekst eller settingen intervensjonen er utført i. Egenskapene som er beskrevet skal være relevante for oversikten informasjonen innhentes til. Eksempler for score på denne komponenten er beskrivelser: av type klasserom, skole, derav privatskole, barneskole, førskole, spesialklasse eller spesialskole. Beskrivelse av fag, lokasjon, kulturelle og sosioøkonomiske faktorer i nærmiljøet, fysisk utforming av omgivelsene studien utføres i.

### **Deltakere**

#### **2.1. Demografiske komponenter**

Studien beskriver informasjon om demografiske egenskaper hos deltakerne som er med i studien. Dette kan eksempelvis være beskrivelser av kjønn, alder, klasstrinn, etnisitet, språkferdigheter, og deltakerne sin sosioøkonomiske status.

#### **2.2. Risikostatus**

Studien gir en beskrivelse av risikostatus eller funksjonshemninger hos deltakerne. Hemninger eksemplifiseres ved læringsvansker, for eksempel lese-, skrive-, stave-, eller hukommelsesvansker som påvirker deltakernes evne til læring, slik som dysleksi. Det kan være autismespekterforstyrrelse, andre kognitive funksjonshemninger, eller atferdsproblematikk. Det er ikke tilstrekkelig ved at klasseleder har sagt at det er en

risikostatus hos deltaker. Det må forekomme beskrivelser av hvilke metoder som har blitt anvendt for å identifisere risikostatus hos deltakerne. For eksempel standardiserte intelligens test, resultater fra nasjonale kartleggingsprøver, *Individuals With Disabilities Education Improvement Act*-kriterium eller vedtak om spesialundervisning.

## **Intervensjonsagent**

### **3.1. Intervensjonsagentens rolle**

Studien beskriver hvilken rolle og egenskaper den som iverksetter intervensjonen har. Det kan være kontaktlærer, forskeren, foreldre, andre familiemedlemmer eller veileder. Dersom det er relevant for artikkelen, skal egenskaper hos intervensjonsagenten beskrives, for eksempel, utdanningsbakgrunn, lisenser og etnisitet.

### **3.2. Opplæring**

Studien beskriver omfanget av trening eller opplæring som intervensjonsagenten har fått i forkant av intervensjonen. Eventuelt hvilke kvalifikasjoner som kreves for å kunne implementere intervensjonen, og hvilke indikasjoner viser at intervensjonsagenten har oppnådd kvalifikasjonene.

## **Beskrivelse av intervensjon**

### **4.1. Intervensjonsprosedyre**

Studien gir fullstendige beskrivelser av kritiske egenskaper ved intervensjonen. Detaljerte beskrivelser av komponenter ved intervensjonen, interesseatferd, kritiske elementer, mengde, og prosedyrer som er skipterte. I tillegg skal intervensjonsagentens handlinger beskrives, for eksempel om det skal utføres prompting, verbale tilbakemeldinger, beskrivelse av fysisk atferd.

### **4.2. Materialer**

Dersom det er relevant med bruk av materiale for gjennomføring av intervensjonen ønskes en beskrivelse av denne informasjonen. Materialer som brukes kan være regneark, tidtakere, signaler og belønninger i form av leker, godteri eller film.

## **Behandlingsintegritet**

### **5.1. Rapportering av behandlingsintegritet**

Studien rapporterer behandlingsintegritet ved bruk av pålitelige og direkte målinger, ofte ved at observatørene tar i bruk en sjekklister for kritiske elementer ved intervensjonen. Dermed måles det i hvilken grad intervensjonen har blitt gjennomført på den måten det var bestemt at den skulle gjennomføres, derav om alle kritiske punkter ble oppfylt.

### **5.2. Rapportering av eksponering**

Studien rapporterer behandlingsintegritet relatert til grad av eksponering og mengde deltakerne ble utsatt for intervensjonen, ved bruk av pålitelige og direkte målinger. Det skal rapporteres om intervensjonens varighet og frekvens, samt om dette har blitt innhentet fra direkte observasjoner eller selv-rapportering fra intervensjonsagenten.

### **5.3. Jevnlig rapportering i alle caser**

Studien gir en vurdering, og rapporterer behandlingsintegritet jevnlig gjennom hele intervensjonsprosedyre, på starten, midten, og slutten av økten, og gjør dette for samtlige av casene studien omhandler. Dersom 5.1 eller 5.2 ikke er beskrevet i studien, kan ikke denne komponenten innvilges (Lane et al., 2014)..

## **Indre validitet**

### **6.1. Kontroll og manipulasjon**

Den uavhengige variabelen er under kontroll av eksperimentet, det betyr at forskere, kontrollerer og systematisk manipulerer uavhengig variabel. Studien beskriver kontroll- og sammenligningsbetingelsene, og de ulike fasene eksperimentet. Et krav for å innvilge komponent 6.1 er at komponent 5.1 om behandlingsintegritet er innvilget (Lane et al., 2014).

## **6.2. Beskrivelse av basislinjen**

Ved single-case studier blir basislinjen beskrevet, eller sammenligningsbetingelsene blir beskrevet ved gruppestudier. For eksempel læreplan, gitte instruksjoner eller andre intervensjoner. Samtidig rapporteres varighet, lengde, frekvens, ratio, og nødvendige definisjoner.

## **6.3. Tilgjengelighet til intervensjonen**

Deltakere i kontrollgruppen, eller deltakerne under basislinjen, har svært begrenset tilgjengelighet til hva intervensjonen innebærer. For eksempel blir det utført måling på basislinjen, før deltakerne får informasjon om hva som skal skje i neste fase, som er tiltaksfasen.

## **6.4. Gruppens oppdrag**

Komponenten er kun aktuelt for gruppedesign, og ble derfor ikke vurdert.

## **6.5. Demonstrasjoner av eksperimentelle effekter**

Det kreves tre demonstrasjoner av eksperimentell effekt, minst tre ganger. Komponenten gjelder kun for studier med single-case design.

## **6.6. Datapunkter i basislinjen**

Det kreves at det fremvises tre datapunkter i alle fasene av basislinjer. Ved fremvisning av færre enn tre datapunkter, må forfatter rettferdiggjøre grunnlaget for det for eksempel ved farlig problematferd uten forbedringspotensialet. Basislinjen må i tillegg vise et etablert og forutsigbart mønster. Komponenten gjelder kun for studier med single-case design, foruten ATD som ikke krever en basislinje-fase.

## **6.7. Trusler mot indre validitet**

Studien har et design som kan utelukke trusler mot indre validitet. Eksempler gitt av slike design er ABAB-design eller MBD med minimum tre deltakere, atferd eller settinger. Komponenter gjelder kun single-case design. Et krav for å innvilge komponent 6.7, er at

komponent 5.1 om behandlingsintegritet, og komponent 6.1 om kontroll og manipulasjon er innvilget (Lane et al., 2014).

### **6.8. Samlet frafall**

Komponenten er kun aktuelt for gruppedesign, og ble derfor ikke vurdert.

### **6.9. Differensiell frafall**

Komponenten er kun aktuelt for gruppedesign, og ble derfor ikke vurdert.

## **Avhengig variabel**

### **7.1. Sosialt validitet**

Studien fremviser et resultat som blir vurdert som sosialt valide. Dette er knyttet til at intervensjonen empirisk eller teoretisk blir begrunnet som viktig for deltakerne sin utvikling og læring, og forbedrer livskvalitet. Den aktuelle studien trenger ikke nødvendigvis å beskrive at resultatet er vurdert som sosialt valide, men intervensjonen studien omhandler skal kunne vurderes som det.

### **7.2. Operasjonalisering av avhengig variabel**

Studien beskriver og definerer målinger på avhengig variabel, og hvilke kriterier som ligger i definisjonen. For eksempel hvilken atferd som er interesseatferden.

### **7.3. Rapportering av effekt**

Alle resultater som forekommer skal bli rapportert i studien, det inkluderer positive, negative og blandet effekter. For eksempel dersom alle caser blir fremstilt i en graf med datapunkter blir komponenten innvilget.

### **7.4. Målinger i alle faser**

Studien fremviser at det er foretatt minimum tre målinger i samtlige faser. I resultatene forekommer det informasjon om frekvens og tid. Færre enn tre målinger per fase i single-case, eller under fire repetisjoner av AB i ATD, krever at forsker har rettferdiggjort grunnlaget for å inkludere færre målinger. For eksempel dersom det foreligger farlig

problematferd uten sannsynlighet for forbedring ved målinger av basislinjen (Lane et al., 2014).

### **7.5. Inter-rater enighet (IOE)**

Studien må ha rapportert og målt IOE, og resultatet må oppnå minimum 80% IOE for at komponenten kan innvilges.

### **7.6. Adekvate bevis for validitet**

Komponenten er kun aktuelt for gruppedesign, og ble derfor ikke vurdert.

## **Dataanalyser**

### **8.1. Tekniker for sammenligning mellom grupper**

Komponenten er kun aktuelt for gruppedesign, og ble derfor ikke vurdert.

### **8.2. Fremstilling av graf**

Studien presenterer grafer som representerer data for alle caser og resultater, dette skal fremvise effekt av intervensjonen. Dette komponenten gjelder for single-case studier. Leser skal enkelt kunne identifisere om det utvises eksperimentell kontroll og konkludere effekt av intervensjonen, etter analysering av fremvist data i grafene. Dette uavhengig om resultatene forekommer som en visuell graf eller på annen måte viser resultater.

### **8.3. Statistikk av effektstørrelse**

Komponenten er kun aktuelt for gruppedesign, og ble derfor ikke vurdert.



## Appendiks B

### Oversikt over treff fra leverandørene OVID, EBSCOhost og Web of Science

**Tabell B1**

*Antall treff i søkedatabasen PsycINFO via OVID*

Søk#	Søkeord	Treff
1	good behavio* game*.mp.	228
2	limit 1 to peer reviewed journal	178
3	*games/	8574
4	game*.mp.	56417
5	3 or 4	56417
6	behavior problems/	31781
7	prosocial behavior/	9054
8	problem behavio* or behavio* problem* or disruptive behavio* or misbehavio* or misconduct or appropriate behavio* or prosocial behavio* or positive behavio* or challenging behavio*).mp.	87388
9	6 or 7 or 8	87388
10	5 and 9	1523
11	classrooms/ or classroom environment/	28312
12	classroom behavior/ or classroom management/	8613
13	classroom behavior modification/	2639
14	school based intervention/	21736
15	elementary schools/	9486
16	exp elementary school students/	47030
17	(school* or classroom*).mp.	559299
18	11 or 12 or 13 or 14 or 15 or 16 or 17	559299
19	10 and 18	364
20	limit 19	140
21	2 or 20	<b>260</b>

*Merknad.* Logg over søkehistorikk for databasen PsycINFO den 17. januar 2023. Søk

nummer 1, 8 og 17 var hovedelementene i problemstillingen og søker etter tekstord, kontra de andre som gjør søk i tesauruser. Søk nummer 2 og 20 viser at søket har blitt avgrenset i databasesøket ved å utelukke det som ikke er fagfellevurderte artikler, samt avgrenset språket til engelsk, dansk, norsk og svensk, og alder til barn 0-12 år og skolealder 6-12 år.

MP=title, abstract, heading word, table of contents, key concepts, original title, tests og measures, mesh word.

**Tabell B2**

*Søketreff via EBSCOhost, i databasene: Education Source, Academic Search Ultimate, ERIC*

Søk#	Søkeord	Education Source	Academic S. Utm	ERIC
1	good behavio* game* OR caught being good game*	134	250	112
2	good behavio* game* OR caught being good game* Limiters – Peer Reviewed	126	236	95
3	DE "Games" OR DE "Educational Games"	12 032	8597	12779
4	game*	138 610	238 456	28,985
5	S3 OR S4	138 610	238 465	28,985
6	DE "Behavior Problems" OR DE "Positive Behavior Supports" OR DE "Prosocial Behavior"	2044	4651	24,824
7	problem behavio* or behavio* problem* or disruptive behavio* or misbehavio* or misconduct or appropriate behavio* or prosocial behavio* or positive behavio* or challenging behavio*	48 614	101 656	39,892
8	S6 OR S7	48 614	101 656	39,892
9	S5 AND S8	853	1958	631
10	S5 AND S8 Limiters – Peer Reviewed	704	1958	409
11	DE "Classrooms" OR DE "Classroom Environment" OR DE "Classroom Techniques" OR DE "Elementary Schools" OR DE "Elementary School Students"	36 326	18 724	108,252
12	school* or classroom* or primary or elementary	1 115 721	5 902 872	988,626
13	S11 AND S12	36 326	18 724	108,252
14	S10 AND S13	27	28	110
15	S10 AND S13 Limiters – Peer reviewed	27	28	110
16	S2 OR S15	138	246	149
17	S2 OR S15 Limiters - Peer reviewed; Language: Danish, English, Norwegian, Swedish	<b>135</b>	<b>241</b>	<b>146</b>

*Merknad.* Logg over søkehistorikken via EBSCO for Education Source, Academic Search

Ultimate og ERIC, den 18.januar 2023. Søk 1, 6 og 11 var de tre elementsøkene, derav er

resterende søk i tesauruser. *Limiters* er avgrensninger i søket, derav kriteriene var

fagfelleverdert artikkel, og språkene engelsk, norsk, dansk og svensk.

**Tabell B3***Søketreff i Web of Science*

Søk	Søkeord	Treff
1	“good behave* game” OR “Caught being good game*” (Topic)	239
2	AND article or Review Article (Document Types) AND English (Languages)	231
3	AND English (Languages)	230
4	game* (Topic) AND "problem behavio*" or "behavio* problem*" or "disruptive behavio*" or "misbehavio*" or "misconduct" or "appropriate behavio*" or "prosocial behavio*" or "positive behavio*" or "challenging behavio*" (Topic) AND school* or classroom* or primary or elementary (Topic)	414
5	AND article or Review Article (Document Types)	413
6	AND English (Languages)	395
7	#3 OR #6	<b>490</b>

*Merknad:* Logg over søkehistorikk utført 18. januar 2023. Denne søkedatabasen har et annet oppsett enn de foregående, der det ble gjort søk på flere temaer samtidig. Søkene 2-3 er begrensninger for søk 1. Søk 4 tar for seg to av hovedelementene for problematferd og for klasserom. Søk 5-6 er begrensninger for søk 4. Begrensninger som var mulig i denne databasen var på artikkeltype og språk.

### **Etisk refleksjonsnotat**

Etiske vurdering har blitt gjennomført fortløpende gjennom skriveprosessen av denne artikkelen. Formålet med en metaanalyse er å samle tidligere studier og sammenfatte resultatene. Ved en slik artikkel, der forfatter selv ikke har utført en egen studie, har det dermed ikke vært nødvendig å være i direkte kontakt med forskningssubjekter. Det har derfor ikke vært aktuelt å ta hensyn til retningslinjer knyttet til personvern og andre formelle forskningsetiske hensyn. Det er ikke nødvendig for ikke-empiriske studier å søke om godkjenning fra forskningsetiske forbund, for eksempel Norsk Senter for Forskningsetikk, som nå er en del av Sikt (2022), eller Regionale Komiteer for medisinsk og helsefaglig forskningsetikk (REK, u.å.). Forfatter har heller ikke måtte gjennomføre en ROS-analyse.

Forfatteren ønsket å se at samtlige studier rapporterte samtykke fra foreldre siden studienes deltakere omhandler mindreårige elever. Et fåtall av studiene oppga dette, men de beskrev at foreldrene, lærer eller skoleledelsen ønsket tiltak i klasserommet grunnet dårlig klasseromsledelse og mye støy. Det var likevel viktig for forfatter å merke seg hvilke studier og databaser forfatter hentet studier fra som skulle inkluderes i artikkelen og metaanalysen. I samarbeid med bibliotekar fra OsloMet ble utvalgt databaser og søkeord bestemt, noe som forfatter vurderte som en trygg kilde til funn av informasjon og studier. Der begrenset vi også innhentede studier til fagfelle-vurderte, som vil si at de har blitt vurdert av fagpersoner innenfor området. Dette kan være en indikator på en legitim studie som ikke krenker andres personvern, og som har opprettholdt de forskningsetiske kravene som foreligger i en slik studie og forskning.

### Referanser

REK (u.å.). *Regionale komiteer for medisinsk og helsefaglig forskningsetikk (REK)*. Hentet

08.05.23 fra <https://rekportalen.no/#hjem/home>

Sikt (01. januar, 2022). *Sikt: Kunnskapssektorens tjenesteleverandør*. Hentet 08.05.23 fra

<https://sikt.no/> org.nr.: 919477822