



UMEÅ UNIVERSITET

Testeffekten och dess relation till arbetsminne, 'need for cognition' och grit

Filip Andersson & Niklas Karlsson

Testeffekten och dess relation till arbetsminne, 'need for cognition' och grit

Forskning i kognitiv psykologi har det senaste decenniet genererat en ansevärd mängd vetenskapliga belägg för att det är mer gynnsamt för framtida hågkomst av ett material om man vid inläringen övar på att plocka fram det genom upprepade test (testbaserad inläring), jämfört med att repetera inläsning av materialet (Carrier & Pashler, 1992; Karpicke & Roediger, 2008; Bertilsson, Wiklund-Hörnqvist, Stenlund, Jönsson, & Jonsson, 2017; se Roediger & Karpicke, 2006a, för en översikt). Detta fenomen kallas för testeffekten (Roediger & Karpicke, 2006a). I en utvärdering av testeffekten redogör Dunlosky, Rawson, Marsh, Nathan och Willingham (2013) för att testeffekten har visat sig vara robust i relation till ett flertal aspekter såsom: olika testformat (flervalsfrågor, essäskrivande, fri återgivning osv.), olika typer av material (från enkelt till mer komplext), olika åldrar (från yngre elever till äldre vuxna), samt representativa utbildningskontexter (autentiska klassrum). Dunlosky och kollegor (2013) framhåller också att testbaserad inläring inte är särskilt tidskrävande eller svårt att implementera i det autentiska klassrummet, men belyser dock, i likhet med Brewer och Unsworth (2012), att forskningen ännu inte vet särskilt mycket om huruvida testeffekten beror på de lärandes individuella egenskaper och färdigheter.

En viktig komponent vad gäller testbaserad inläring är återkoppling (feedback). Återkoppling i form av korrekt svar efter varje framlockning blir viktig eftersom det möjliggör ytterligare ett inkodningstillfälle samt motverkar felinläring (Wiklund-Hörnqvist, Andersson, Jonsson, & Nyberg, 2017). Testbaserad inläring med återkoppling påvisar även en större testeffekt jämfört med då man inte får återkoppling (Butler, Karpicke, & Roediger, 2008; Shute, 2008). Om syftet är att använda testbaserad inläring för att stärka lärandet bör återkoppling inkluderas.

I en metaanalys visade Rowland (2014) att testeffekten är robust både för längre tidsintervall (en dag eller mer efter inläring) och för kortare tidsintervall (mindre än en dag efter inläring), om än i mindre grad. I studier där hågkomsttest har använts vid korta tidsintervall (några minuter efter inläring) har dock motstridiga resultat observerats (Rowland, 2014). Vissa studier har visat att testbaserad inläring resulterar i bättre hågkomst än läsning direkt efter inläring (Bertilsson m. fl., exp. 2, 2017; Carpenter & DeLosh, 2005, 2006; Karpicke & Zaromb, 2010; Kuo & Hirshman, 1996; Rowland & DeLosh, 2014). Andra studier har visat att det inte är någon skillnad eller att läsning resulterar i bättre hågkomst än testbaserad inläring direkt efter inläring (t. ex. Bertilsson m. fl., exp. 1, 2017; Roediger & Karpicke 2006b; Wheeler, Ewers, & Buonanno, 2003; Toppino & Cohen, 2009; Congleton & Rajaram, 2012). Det är dock få studier med inomgruppsdesign som har undersökt hågkomst av material vid olika tidsintervall, där deltagarna inte har exponerats för inlärningsmaterialet sedan inlärningsfasen, eftersom de flesta studier återanvänder samma material i samtliga hågkomsttest (Stenlund, Jönsson & Jonsson, 2017) och/eller nyttjar en mellangruppsdesign (Wiklund-Hörnqvist, Jonsson & Nyberg, 2014). I en tidigare studie av Bertilsson och kollegor (2017; experiment 2) där man jämförde effekter av testbaserad inläring och läsning direkt efter inlärningsfasen, en vecka efter, samt fyra veckor efter inläring, testades både inlärningsmaterial som deltagarna inte har exponerats för sedan inlärningsfasen (unik

material), och material som var med i föregående hågkomsttest (som tillsammans med det unika materialet utgör ackumulerat material). I experiment 2 där man använde en inomgruppsdesign fann Bertilsson och kollegor (2017) en fördel avseende hågkomst för testbaserad inläring jämfört med läsning vid samtliga hågkomsttest, både för unika och ackumulerade ordpar.

Trots att det finns evidens för testeffekten så föreligger det fortfarande ett fåtal studier som har undersökt effekterna av testbaserat lärande i relation till individuella variationer avseende arbetsminneskapacitet (AMK) samt personlighetsdrag som är relevanta för inläring (Dunlosky & kollegor, 2013). Om testbaserad inläring ska rekommenderas som en god pedagogisk metod är det viktigt att undersöka potentiella faktorer som har visats påverka inlärningsförmåga.

Relationen mellan inläring och individuella variationer avseende AMK har påvisats i en ansevärd mängd studier som bland annat berör skrivförmåga (Swanson & Berninger, 1996), matematisk kompetens (Smedt m. fl., 2009) och förmågan att lära sig ett andraspråk (Juffs & Harrington, 2011). Arbetsminne kan beskrivas som ett minnessystem som möjliggör temporär lagring och manipulering av information (Baddeley, 1992). Individuella variationer avseende AMK har visat sig vara prediktiva både för inläring (Bergman-Nutley & Klingberg, 2014) och skolprestationer (Alloway & Alloway, 2010; Wiklund-Hörnqvist, Jonsson, Korhonen, Eklöf, & Nyroos, 2016), och man har även funnit evidens för att personer med lägre AMK har visat sig ha problem med att plocka fram information från långtidsminnet (Unsworth, Brewer, & Stillers, 2013).

Något som också har visat sig spela roll för grad av inläring är 'need for cognition' (NFC), vilket handlar om engagemang för samt åtnjutande av tänkande och problemlösning (Cacioppo & Petty, 1982; Dornic, Ekehammar, & Laaksonen, 1991). Individuella variationer avseende NFC har påvisat ett positivt samband med prestation på matematiska problemlösningsuppgifter (Dornic, Ekehammar, & Laaksonen, 1991). NFC har även visat sig predicera vilka studenter som delas in i klasser med högre nivåer av begåvning (Meier, Vogl, & Preckel, 2014). Därtill har Stenlund, Jönsson, & Jonsson (2017) visat att gymnasieelever med högt NFC hade bättre hågkomst än de med lågt NFC när de hade lärt in materialet genom gruppdiskussioner. Forskning har även funnit indikationer på att lärande påverkas av individuella variationer avseende grit - vilket har definierats som "ihärdighet och passion för långsiktiga mål" (Duckworth, Peterson, Matthews, & Kelly, 2007, s. 1087). Wolters och Hussain (2015) fann att grit hade betydelse för förmågan till självreglerande lärande, vilket de beskriver som ett aktivt och målmedvetet sätt att reglera motivationsmässiga, kognitiva och beteendenaspekter av sitt egna lärande. Bertilsson och kollegor (2017) benämner både NFC och grit som personlighetsdrag som påverkar motivationen att lära.

Ett fåtal testeffektsstudier har undersökt hur AMK (Bertilsson m. fl., 2017; Wiklund-Hörnqvist, Jonsson, & Nyberg, 2014), NFC (Bertilsson m. fl., 2017; Stenlund, Jönsson, & Jonsson, 2017) och grit (Bertilsson m. fl., 2017) påverkar hågkomst av material som lärts in genom läsning samt genom testbaserad inläring. Varken Wiklund-Hörnqvist och kollegor (2014) eller Bertilsson och kollegor (2017) fann någon inverkan av individuella variationer avseende AMK på hågkomst vid något av hågkomsttesten (varken direkt efter

inlärnin g, en vecka efter eller fyra veckor efter) varken av ord som hade lärts in med läsning eller testbaserad inlärnin g med återkoppling. Bertilsson och kollegor (2017) fann heller ingen effekt av NFC eller grit för grad av hågkomst. Stenlund, Jönsson och Jonsson (2017) använde en design med samma tidsintervall mellan hågkomsttesten som Bertilsson och kollegor (2017) och fann inte heller någon signifikant inverkan av NFC för senare hågkomst för deltagare som hade lärt in materialet genom läsning eller testbaserad inlärnin g med återkoppling.

De ovan nämnda studierna undersöker alla hur individuella variationer avseende AMK, NFC och/eller grit relaterar till hågkomst efter testbaserad inlärnin g samt efter läsning. Det är emellertid inte så många studier som har undersökt hur individuella variationer avseende AMK, NFC och/eller grit relaterar till skillnaden mellan hågkomst efter testbaserad inlärnin g och läsning (definierat som testeffektens magnitud; Brewer & Unsworth, 2012). För att få insikt om huruvida individuella variationer påverkar hur mycket bättre hågkomst testbaserad inlärnin g genererar jämfört med läsning bör man undersöka testeffektens magnitud, vilket förutsätter en inomgruppsdesign. Frågan om huruvida individuella variationer avseende NFC och grit kan relateras till testeffektens magnitud verkar fortfarande vara outforskad. Vad gäller AMK och testeffektens magnitud har ett fåtal studier genomförts. Varken Brewer och Unsworth (2012) eller Tse och Pu (2012) fann någon signifikant relation mellan individuella skillnader i AMK och testeffektens magnitud. Däremot fann Agarwal, Finley, Rose, & Roediger (2016) att testeffektens magnitud minskade som funktion av högre AMK vid ett hågkomsttest som gjordes två dagar efter inlärnin gen, det vill säga att de som hade lägre AMK gynnades mer av testbaserad inlärnin g med återkoppling jämfört med de som hade högre AMK. Detta resultat avsåg de frågor som deltagarna fick återkoppling på när de testades två dagar tidigare, men ingen signifikant korrelation identifierades avseende de frågor som deltagarna inte fick återkoppling på (Agarwal m. fl., 2016). Forskningen inom detta område ligger emellertid fortfarande i sin linda, och fler studier behövs för att kunna påvisa vetenskaplig evidens för om, och hur, de gynnsamma effekterna av testbaserad inlärnin g med återkoppling influeras av individuella variationer avseende AMK, NFC och grit.

Den föreliggande studiens syften är att: 1) undersöka testeffektens robusthet för olika tidsintervall både för unika och för ackumulerade ordpar, samt 2) undersöka huruvida individuella variationer avseende AMK, NFC och grit relaterar till testeffektens magnitud hos gymnasieelever.

Hypoteser

1. *AMK påverkar inte testeffektens magnitud.* Det skulle kunna vara så att testning jämnar ut arbetsminnets effekter på inlärnin g, så att lågt AMK innebär större nytta av testeffekten, vilket skulle förklara att Agarwal och kollegor (2016) identifierade detta samband. Brewer och Unsworth (2012) för fram resonemanget att personer med hög kognitiv förmåga möjligen utnyttjar sina kognitiva förmågor i högre grad jämfört med personer med lägre kognitiv förmåga oavsett vilken inlärnin gsmetod de använder. Om

det stämmer skulle det kunna vara så att personer med hög AMK inte har lika stor fördel av inlärningsmetoder som underlättar inläringen. Brewer och Unsworth (2012) framför även ett alternativt resonemang som föreslår att testbaserad inläring möjliggör utnyttjandet av ens inneboende förmågor i högre utsträckning än läsning. Eftersom hög AMK innebär att det finns mer kognitiv förmåga att utnyttja, borde detta resonemang innebära att personer med hög AMK har större nytta av testeffekten. Men eftersom de flesta av de studier som är jämförbara med den föreliggande studien har funnit att AMK inte har en signifikant inverkan på testeffektens magnitud (t. ex. Brewer & Unsworth, 2017) är hypotesen att testning ger likvärdiga fördelar för personer oberoende av AMK.

2. *NFC påverkar testeffektens magnitud.* Eftersom testbaserad inläring involverar läsning, framplockning, samt erhållande av återkoppling, medan läsbetingelsen endast involverar läsning, är det rimligt att anta att testbetingelsen kräver mer tankeverksamhet. NFC handlar om engagemang för, samt åtnjutande av tänkande och problemlösning, vilket gör det rimligt att anta att personer med högt NFC blir mer engagerade i uppgifter som kräver mer tänkande och problemlösning. Om dessa två antaganden stämmer kan personer med högt NFC antas få större fördelar i form av ökat engagemang vid testbaserad inläring jämfört med läsning.
3. *Grit påverkar inte testeffektens magnitud.* Det är tänkbart att hög grit kan bidra till bättre resultat i testbetingelsen jämfört med låg grit. Grit handlar om ihärdighet och passion för mål, vilket skulle kunna göra att deltagare med hög grit inte blir lika nedslagna efter ett misslyckat framplockningsförsök, utan bibehåller sin motivation och anstränger sig för att lyckas bättre vid nästa framplockningstillfälle. Grit avser dock endast långsiktiga mål, medan experimentet i denna studie pågår under ett kortare tidsspann (fyra veckor). Därför är det inte troligt att grit påverkar testeffektens magnitud.

Metod

Deltagare

Urvalet i studien bestod av 182 gymnasieelever från en gymnasieskola i norra delen av Sverige och kom från följande gymnasieinriktningar: naturvetenskap; teknik; barn och fritid; vård och omsorg. Tolv deltagare exkluderades för att de inte deltog i studiens alla moment och ytterligare fem deltagare exkluderades eftersom deras resultat skiljde sig åt med mer än ± 3 standardavvikelse från medelvärdet. Det slutgiltiga urvalet bestod av 165 deltagare (flickor, $n = 81$). Alla deltagare var mellan 16 och 20 år, och genomsnittsåldern var 17.14 ($SD = .621$). Samtliga deltagare erhöll både skriftlig och muntlig information om studien innan datainsamling påbörjades. Informerat samtycke erhöles skriftligen från samtliga deltagare i enlighet med Helsingforsdeklarationen och studien hade godkänts av regionala etikprövningsnämnden vid Umeå universitet (2015/382-31Ö). Alla moment i studien utfördes

under lektionstid. Varje deltagare fick biobiljetter som ersättning. En biobiljett erhöles per moment.

Instrument och material

Inlärningsmaterial. Materialet som användes för att mäta testeffekten utgjordes av 60 stycken swahili-svenska ordpar (Karlsson Wirebring m. fl., 2015; Nelson & Dunlosky, 1994). Ingen av deltagarna rapporterade att de kunde swahili sedan tidigare. Resultatet anges i andel korrekt återgivna ordpar (antal korrekt återgivna ordpar / totala antalet ordpar: se design och procedur).

Mätningar av kognitiva förmågor och personlighetsdrag. För att undersöka i vilken mån individuella skillnader i kognitiva förmågor och aspekter av personlighetsdrag påverkar magnituden av testeffekten gjordes ett flertal mätningar av deltagarnas kognitiva kapaciteter och personliga egenskaper inom ramen för ett större projekt. Endast mätningarna av AMK, NFC och grit används i denna studie.

Arbetsminneskapacitet (AMK). För att mäta AMK användes ett datorbaserat komplext arbetsminnestest: Automatic version of Operation Span task (Aospan; Unsworth m. fl., 2005). I ett Aospan-test ska deltagaren lösa ett antal matematiska problem samtidigt som en sekvens av bokstäver ska memoreras. Testet består av fem olika set-storlekar innehållande tre, fyra, fem, sex respektive sju komponenter. En komponent består av en ekvation som deltagaren ska bedöma som sann eller falsk, efterföljt av en presenterad bokstav. Varje set-storlek är med tre gånger vardera, vilket innebär att testet innehåller sammanlagt 75 matematikproblem och 75 bokstäver. Efter varje set skulle deltagarna ange de presenterade bokstäverna i rätt ordning. Antalet korrekt återgivna bokstäver, givet att de återgavs i korrekt ordning, användes i den statistiska analysen (Agarwal m. fl., 2016; Bertilsson m. fl., 2017). Aospan har påvisat god intern konsistens, .78, och god test-retest reliabilitet på .83 (Unsworth m. fl., 2005).

Need for cognition (NFC). För att mäta NFC användes en förkortad variant av Mental Effort Tolerance Questionnaire (METQ; Dornic m.fl, 1991), vilket är en svensk version av NFC skalan (Cacioppo & Petty, 1982). Den förkortade varianten av METQ har enligt Dornic m. fl. (1991) en intern konsistens på .90. Formuläret utgörs av påståenden där svaren anges i en 5-gradig skala där 1 = instämmer absolut inte och 5 = instämmer helt. Summan av svaren utgör resultatet på formuläret, vilket betraktas som deltagarens NFC. I det ursprungliga METQ-formuläret reflekterade hälften av påståendena positiva attityder och den andra hälften negativa attityder kring åtnjutande av tänkande. Den förkortade versionen som användes i den föreliggande studien innehåller tolv positivt formulerade påståenden och 18 negativt formulerade påståenden (se Stenlund & Jonsson, 2017 för mer detaljer). Svaren på de negativt formulerade påståendena räknas omvänt.

Grit. För att mäta grit användes Short Grit Scale (Grit-S). Grit-S är ett formulär som innehåller åtta påståenden och är en omarbetad version av den ursprungliga Grit Scale (Grit-O). Grit-S är kortare samt har bättre psykometriska egenskaper än Grit-O och betraktas därför som ett bättre alternativ. Duckworth & Quinn (2009) har rapporterat att Grit-S har en

intern konsistens mellan 0.73 och .84. Frågeformuläret som användes i den föreliggande studien översattes från engelska till svenska, för att sedan översättas tillbaka till engelska av en oberoende professionell översättare, och det visade sig att resultatet i väldigt hög grad liknade det ursprungliga formuläret (Bertilsson m. fl., 2017). Formuläret utgörs av påståenden där svaren anges i en 5-gradig skala, från 1 = *stämmer inte alls* till 5 = *stämmer helt och hållet*. Summan av svaren utgör resultatet på formuläret, vilket betraktas som deltagarens grit (Bertilsson m. fl., 2017). Svaren på de negativt formulerade påståendena räknas omvänt.

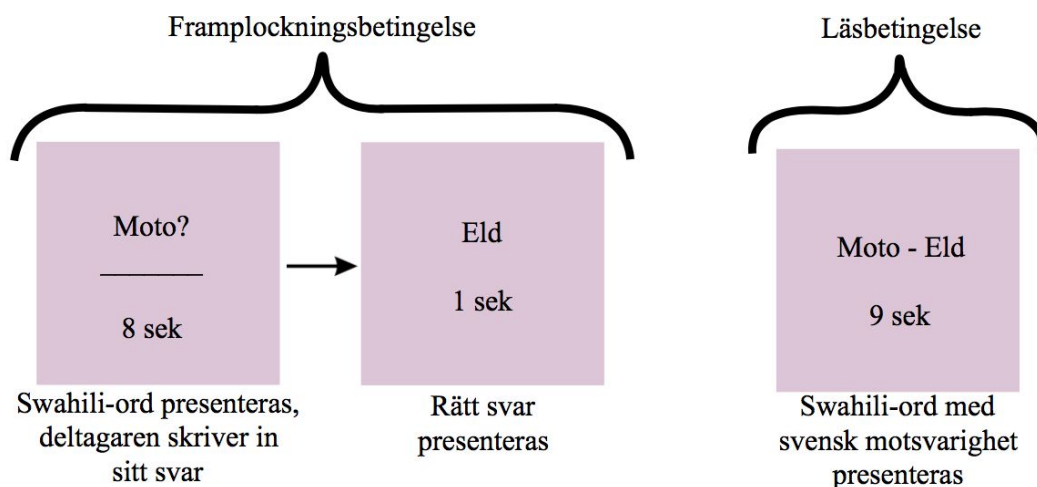
Design och procedur

Den föreliggande studien använde en inomgruppsdesign i syfte att undersöka lärandeffekter av två olika inlärningsmetoder (läsning och testbaserat lärande) som samtliga deltagare använde för att lära in swahili-svenska ordpar (se Figur 1).

En vecka före interventionen fyllde samtliga deltagare i formulären som behandlade NFC och grit med papper och penna, och sedan genomförde de ett datoriserat testbatteri avseende AMK och ytterligare några test av kognitiva förmågor som inte användes i den föreliggande studien. Försökspersonerna tog del av studien under lektionstid, och för alla moment som genomfördes med hjälp av dator försågs nödvändig utrustning av skolan.

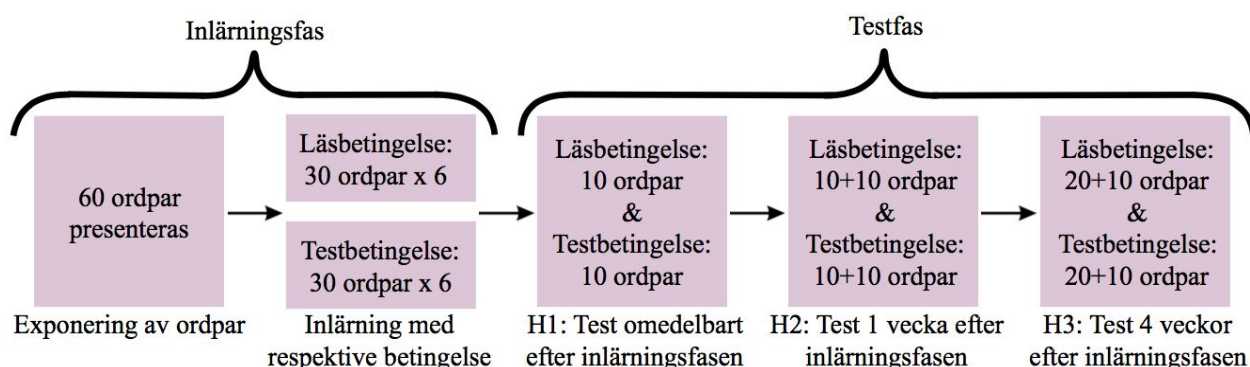
Experimentet började med en inlärningsfas som inleddes med att samtliga deltagare presenterades för alla 60 swahili-svenska ordpar på datorskärmen en gång [Moto - Eld]. Varje ordpar visades ett i taget under åtta sekunder. Därefter påbörjades interventionen; hälften av ordparen lärdes in med läsbetingelsen medan den andra hälften av ordparen lärdes in genom testbaserad inläring med återkoppling (se Figur 1). Detta resulterade i 30 ordpar per betingelse. Varje ordpar repeterades sex gånger. Vilka ordpar som tillhörde vilken betingelse randomiserades för varje deltagare. Inom varje repetition randomiserades ordparens ordning med kontroll för att ett och samma ordpar inte hamnade för nära sig självt i efterföljande omgångar.

I läsbetingelsen blev deltagarna instruerade att läsa ett ordpar som presenterades i nio sekunder. Ordparet bestod av ett swahiliord till vänster och dess svenska översättning till höger [Moto - Eld]. I testbetingelsen blev deltagarna instruerade att skriva in den svenska översättningen av ett swahiliord som visades på skärmen följt av ett frågetecken [Moto?]. Deltagarna hade åtta sekunder på sig att läsa swahiliordet och skriva in dess svenska översättning, varefter de fick återkoppling i form av det korrekta svenska ordet som presenterades i en sekund. Deltagarna fick denna återkoppling oavsett om de hade svarat rätt, fel eller inte angivit något svar. Se figur 1 för en överskådlig genomgång av de två betingelserna.



Figur 1. Läs- och testbetingelsen. I testbetingelsen presenteras först ett ord på swahili samtidigt som deltagaren har möjlighet att fylla i den svenska motsvarigheten till det presenterade swahili-ordet under åtta sekunder, vilket efterföljs av att återkoppling i form av rätt svar presenteras i en sekund. I läsbetingelsen presenteras swahiliordet tillsammans med den svenska motsvarigheten under nio sekunder.

Inlärningsfasen efterföljdes av en fem minuter lång distraktionsövning i form av ett datorspel som programmet automatiskt länkade till efter avslutad intervention. Efter fem minuter av spelande påbörjade deltagarna det första hågkomsttestet (H1). H1 bestod av 10 ordpar från testbetingelsen och 10 ordpar från läsbetingelsen. Ett till hågkomsttest utfördes en vecka senare (H2), och ytterligare ett hågkomsttest utfördes fyra veckor senare (H3). H2 bestod av de 20 ordpar som ingick i H1 samt 20 unika ordpar (10 från testbetingelsen och 10 från läsbetingelsen) som deltagarna inte hade exponerats för sedan inlärningsfasen en vecka tidigare. I H2 testades på så sätt 20 ordpar per betingelse, och följaktligen 40 ordpar sammanlagt (kallas här efter för ackumulerade ordpar). På samma sätt bestod H3 av de 40 ordpar som ingick i H2 samt 20 unika ordpar (10 från testbetingelsen och 10 från läsbetingelsen) som deltagarna inte hade exponerats för sedan inlärningsfasen fyra veckor tidigare. I H3 testades på så sätt 30 ackumulerade ordpar per betingelse, och följaktligen 60 ackumulerade ordpar (se Figur 2 för ett schematiskt flöde av experimentet).



Figur 2. Studiens design. Under inlärningsfasen får deltagarna initialt 60 ordpar presenterade. Därefter presenteras deltagarna för hälften av ordparen i sin helhet (läsbetingelsen) medan de exponeras för den andra hälften av ordparen genom testbaserad inlärnin g med återkoppling (testbetingelsen). Under den efterföljande testfasen testas deltagarna på ordparen. Omedelbart efter inlärningsfasen testades deltagarna på 10+10 ordpar från vardera betingelse (H1), en vecka senare testas de på de 20 ordpar som ingick i det föregående testet samt 10+10 nya ordpar från vardera betingelse (H2), och fyra veckor senare testades de på de 40 ordpar som ingick i de föregående testerna samt 10+10 nya ordpar från vardera betingelse (H3).

Statistiska analyser

Analyserna i den föreliggande studien använder inlärningsbetingelserna (läsning och testbaserad inlärnin g med återkoppling), tidsintervallen mellan inlärnin g och hågkomsttest (direkt efter, en vecka efter och fyra veckor efter), och AMK, NFC och grit som oberoende variabler. De beroende variablerna är resultatet på hågkomsttesten (H1, H2 och H3) samt testeffektens magnitud på dessa hågkomsttest. Resultaten på hågkomsttesten är andel korrekta svar för respektive inlärningsbetingelse (antal korrekt återgivna ordpar / totala antalet ordpar). Testeffektens magnitud mäts genom att beräkna differensen i andelen korrekta svar på respektive hågkomsttest mellan de två betingelserna (andelen korrekta svar på orden från testbetingelsen minus andelen korrekta svar på orden från läsbetingelsen; Brewer & Unsworth, 2012; Agarwal, 2016).

Eftersom det för bägge betingelserna tillkom tio nya ordpar vid varje hågkomsttest (H1, H2 respektive H3) analyserades testeffekten både för de tio unika ordparen och de ackumulerade ordparen i två separata 2 x 3 upprepade mätningar MANOVA-analyser. Den ena MANOVA-analysen undersökte de unika ordparen och den andra undersökte de ackumulerade ordparen. Detta innebär att analyserna beaktar tio unika ordpar per betingelse för samtliga hågkomsttest (H1, H2 och H3), samt 20 ackumulerade per betingelse för H2 och 30 ackumulerade per betingelse för H3. På så vis undersöktes hågkomstens relation till både inlärningsbetingelse och tidsintervall mellan inlärningsfas och hågkomsttest. De tio unika ordparen per betingelse från H1 ingick i bägge analyserna.

Mauchly's sfäricitetstest utfördes för bägge MANOVA-analyserna och i de fall detta indikerade att antagandet om sfäricitet hade överträtts valdes Greenhouse-Geisser-estimatet som korrigering ifall antingen detta estimat eller Huynh-Feldt-estimatet var under .075, och i annat fall valdes Huynh-Feldt-estimatet (Huynh & Feldt, 1976; Howell, 2002). I bägge MANOVA-analyserna anges effektstorleken för F -värden i partial eta squared (η_p^2) och samtliga signifikansprövningar korrigerades enligt Bonferroni-metoden (Field, 2009).

Med grund i tidigare studier som har visat att AMK (Swanson & Berninger, 1996; Smedt m. fl., 2009; Juffs & Harrington, 2011), NFC (Dornic, Ekehammar, & Laaksonen, 1991; Meier, Vogl, & Preckel, 2014) och grit (Wolters & Hussain, 2015) har betydelse för lärande, gjordes en simultan multipel regressionsanalys som undersökte hur AMK, NFC och grit relaterar till testeffektens magnitud, utan föregående korrelationsanalys. En av anledningarna till att korrelationsanalys uteslöts var att undvika förhöjd risk för typ II-fel som

kan bli konsekvensen av att korrigera för multipla jämförelser, när det är ett högt antal signifikansprövningar som behöver korrigeras (Field, 2009).

Inför utförandet av regressionsanalysen testades de relevanta antagandena (normalitet, linjäritet, homoskedasticitet och multikollinearitet) i enlighet med vad som föreslås av Field (2009). Dessa antaganden bedömdes gälla, varför resultatet både kan antas gälla för urvalet och vara generaliserbart för populationen (Field, 2009).

I tabellen över regressionsanalysens resultat redovisas okorrigerade p -värden med asterisk bredvid om p -värdet var signifikant efter korrigering av alpha enligt Holm-Bonferroni-metoden ($\alpha / (n - i + 1)$) där n är antalet signifikansprövningar och i är respektive signifikansvärdes rangordning i förhållande till de andra signifikansvärdena (lägst p -värde = 1, näst lägst p -värde = 2 osv.; Holm, 1979; Aickin & Gensler, 1996). När denna metod har använts har $\alpha = .05$ använts som utgångsvärde. Antalet signifikansprövningar har utgått från hur många prediktorers inverkan på respektive beroende variabel som har signifikansprövats.

Fem deltagare uteslöts från analysen på grund av att de befann sig ± 3 standardavvikelser från medelvärdet för någon av de analyserade variablerna. Ytterligare sju deltagare uteslöts från analyserna som involverade AMK, NFC och grit, eftersom dessa deltagare inte hade genomfört testerna som mäter dessa konstrukt. Alla variabler var normalfördelade med skevhetsvärden mellan -.934 och 1.277 och kurtosisvärden mellan -.868 och .676 (Finney & DiStefano, 2006; Curran, West och Finch, 1996). En envägs ANOVA-analys uppvisade ingen effekt av kön på något av hågkomsttesten eller för magnituden av testeffekten (alla $p > .129$), vilket medförde att kön som variabel uteslöts från samtliga statistiska analyser.

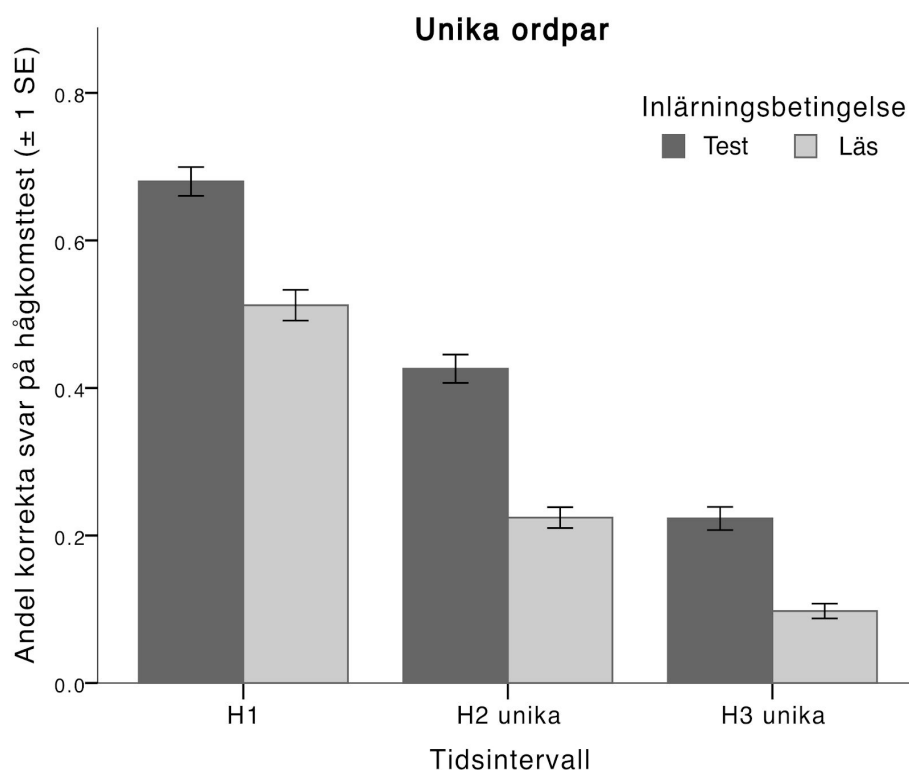
Resultat

Nedan presenteras resultatet från två stycken upprepade mätningar MANOVA analyser (unika respektive ackumulerade) som ämnar undersöka huruvida testeffekten bekräftas både för unika och ackumulerade ordpar vid olika tidpunkter efter inlärningsfasen. Därefter följer en regressionsanalys avseende testeffektens magnitud i relation till individuella variationer avseende AMK, NFC och grit för respektive tidpunkt.

Testeffekten och unika ordpar.

En 2 (inläring: läs- & testbetingelse) x 3 (tid: H1, H2, & H3) upprepade mätningar MANOVA påvisade en signifikant huvudeffekt av inlärningsbetingelse på hågkomst $F(1, 164) = 284.543$, $MSE = .024$, $p < .001$, $\eta_p^2 = .634$, vilket indikerar att hågkomsten är bättre efter testbaserad inläring jämfört med efter läsning oberoende av tidpunkt (se Figur 3). Analysen fann även en huvudeffekt av tid, $F(1.850, 303.430) = 474.886$, $p < .001$, $\eta_p^2 = .743$ vilket indikerar att oberoende av inläring så glömmar man över tid. Vidare identifierades en signifikant interaktionseffekt mellan inlärningsbetingelse och tidsintervall $F(2, 328) = 6.564$, $MSE = .018$, $p = .002$, $\eta_p^2 = .038$. Parvisa kontraster påvisade att denna interaktionseffekt drevs av att deltagarnas procentuella glömska av ordparen från läsbetingelsen var lägre

mellan H2 och H3 jämfört med mellan H1 och H2, medan förändringen inte var lika stor för ordparen från testbetingelsen. Det bör dock nämnas att den procentuella glömskan [(initial hågkomst – senare hågkomst) / initial hågkomst] var högre för ordparen i läsbetingelsen jämfört med testbetingelsen både mellan H1 och H2 (läsbetingelse: 56,25 %; testbetingelse: 37,35 %), och mellan H2 och H3 (läsbetingelse: 56,25 %; testbetingelse: 47,65 %).

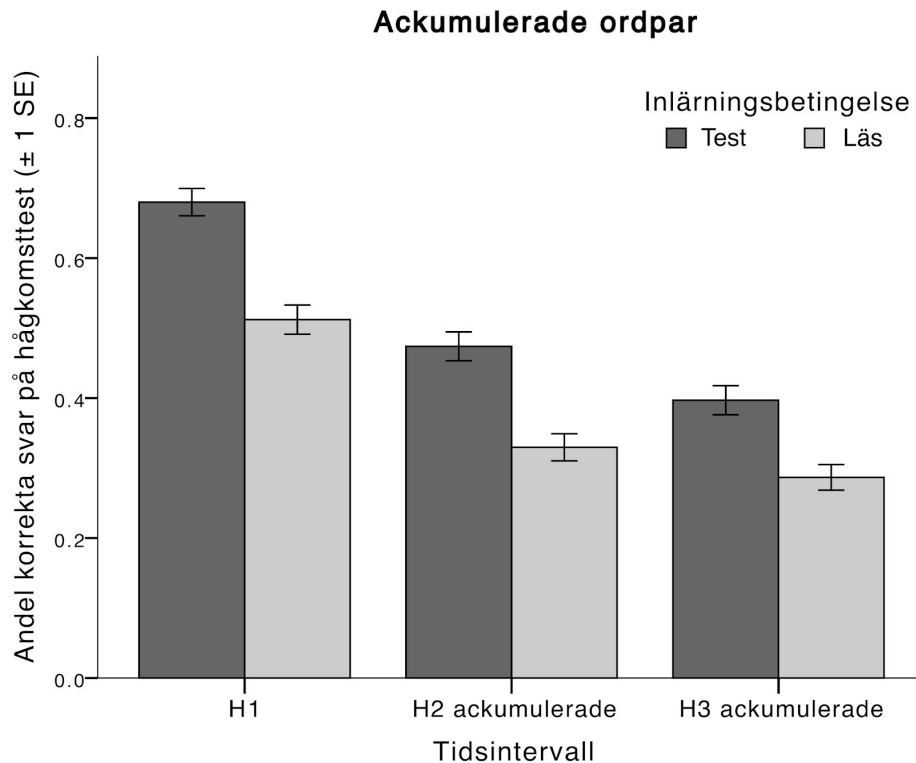


Figur 3. Andel korrekta svar på de unika ordparen för läs- respektive testbetingelsen vid H1 (direkttest), H2 (en vecka efter) och H3 (fyra veckor efter). Felstaplar representerar ± 1 standardfel.

Testeffekten och ackumulerade ordpar.

En 2 (inläring: läs- & testbetingelse) x 3 (tid: H1, H2, & H3) upprepade mätningar MANOVA påvisade en signifikant huvudeffekt av inlärningsbetingelse på hågkomst $F(1, 164) = 78.327$, $MSE = .063$, $p < .001$, $\eta_p^2 = .323$, vilket indikerar att oberoende av tidpunkt är hågkomsten bättre efter testbaserad inläring jämfört med efter läsning (se Figur 4). Analysen fann även en huvudeffekt av tid på hågkomst $F(1.606, 263.394) = 274.798$, $MSE = .026$, $p < .001$, $\eta_p^2 = .626$, vilket indikerar att oberoende av inläring så glömmes man över tid. Vidare identifierades en signifikant interaktionseffekt mellan inlärningsbetingelse och tidsintervall, $F(1.697, 278.342) = 5.204$, $MSE = .016$, $p = .009$, $\eta_p^2 = .031$. Parvisa kontraster påvisade att denna interaktionseffekt drevs av att deltagarnas procentuella glömska av ordparen från läsbetingelsen var lägre mellan H2 och H3 jämfört med mellan H1 och H2, medan förändringen inte var lika stor för ordparen från testbetingelsen. Den procentuella glömskan

$[(\text{initial h\AA}g\text{komst} - \text{senare h\AA}g\text{komst}) / \text{initial h\AA}g\text{komst}]$ var h\AA}gre f\AA}r ordparen i l\AA}sbetingelsen j\AA}mf\AA}rt med testbetingelsen mellan H1 och H2 (l\AA}sbetingelse: 35.55 %; testbetingelse: 30.29 %), medan det motsatta f\AA}rh\AA}llandet observerades mellan H2 och H3 (l\AA}sbetingelse: 13.03 %; testbetingelse: 16.88 %).



Figur 4. Andel korrekta svar p\AA} de unika ordparen f\AA}r l\AA}s- respektive testbetingelsen vid H1 (direkttest), H2 (en vecka efter) och H3 (fyra veckor efter). Felstaplar representerar ± 1 standardfel.

Inverkan av individuella variationer avseende AMK, NFC och grit f\AA}r testeffektens magnitud.

F\AA}r att unders\AA}ka om, och i s\AA} fall hur AMK, NFC och grit relaterar till testeffektens magnitud genomf\AA}rdes en regressionsanalys i vilken resultatet p\AA} Ospan, METQ, och grit anv\AA}ndes som prediktorer f\AA}r testeffektens magnitud f\AA}r samtliga h\AA}gkomsttest, och f\AA}r unika och ackumulerade ordpar var f\AA}r sig (se Tabell 1). Analysen visade att ingen av prediktorerna hade signifikant inverkan p\AA} magnituden av testeffekterna efter att α -v\AA}rdet korrigerats med Holm-Bonferroni-metoden (Holm, 1979; Aickin & Gensler, 1996).

Tabell 1

Översikt av multivariat simultan regressionsanalys predicera testeffektens magnitud med hjälp av poäng på Ospan, METQ och Grit, för varje hågkomsttest och för både unika och ackumulerade ordpar ($N = 158$)

Testeffektens magnitud Testtillfälle / Ordpar	Prediktorer	<i>B</i>	<i>SE B</i>	β	<i>t</i>	<i>p</i>
H1	AMK	.001	.001	.58	.696	.488
	NFC	-.001	.001	-.073	-.843	.401
	Grit	-.002	.036	-.006	-.067	.946
H2 Unika	AMK	.002	.001	.167	2.035	.044
	NFC	.001	.01	.109	1.289	.199
	Grit	.001	.029	.003	.040	.968
H2 Ackumulerade	AMK	.002	.001	.137	-1.313	.100
	NFC	.000	.001	-.019	-.220	.826
	Grit	-.051	.039	-.108	1.654	.191
H3 Unika	AMK	.000	.001	-.030	-.359	.720
	NFC	.000	.001	.008	.097	.923
	Grit	.038	.032	.098	1.186	.237
H3 Ackumulerade	AMK	.002	.002	.093	1.112	.268
	NFC	.000	.001	.012	.142	.887
	Grit	-.024	.044	-.045	-.545	.587

Kommentar: Ospan (operation span task) = mått på arbetsminneskapaciteten. METQ (Mental Effort Tolerance Questionnaire) = need for cognition. Grit = mått på ihärdighet och vilja att uppnå sina mål. Variablerna på den vertikala axeln betecknar testeffektens magnitud för de olika hågkomsttesten (H) och efterföljande siffra betecknar vilket test det rör sig om (1 = hågkomsttestet direkt efter inläring, 2 = en vecka efter och 3 = 4 veckor efter). Unika betecknar att variabeln endast avser de nya ordpar som tillkom för varje hågkomsttest, medan ackumulerade betecknar att variabeln avser både de ordpar som testades av föregående hågkomsttest och 10 nya för båda betingelserna.

I analysen ingick även en regressionsmodell som visade huruvida AMK, NFC och grit tillsammans har någon inverkan på testeffektens magnitud på de tre hågkomsttesten för unika respektive ackumulerade ordpar: H1 [$R^2 = .007$, $F(3,154) = .338$, $p = .798$]; H2 unika [$R^2 = .05$, $F(3,154) = 2.724$, $p = .046$]; H3 unika [$R^2 = .011$, $F(3,154) = .564$, $p = .639$]; H2 ackumulerade [$R^2 = .030$, $F(3,154) = 1.601$, $p = .191$]; och H3 ackumulerade [$R^2 = .011$, $F(3,154) = .584$, $p = .627$]. Ingen av modellerna var signifikanta efter att α -värdena korrigerats med Holm-Bonferroni-metoden. Sammantaget indikerar resultatet från regressionsanalyserna att prediktorerna - varken var för sig eller tillsammans - hade någon

inverkan på testeffektens magnitud. Detta indikerar att testeffektens storlek är jämlig bland individer som skiljer sig åt avseende AMK, NFC och grit.

Diskussion

Ett av den föreliggande studiens syften var att undersöka testeffektens robusthet för olika tidsintervall, samt för unika och ackumulerade ordpar. Resultatet påvisade en signifikant testeffekt både för unika och ackumulerade ordpar och adderar till den växande mängden vetenskapliga belägg för att testbaserad inläring är överlägsen inläring genom läsning avseende hågkomst (t. ex. Roediger & Karpicke, 2006a). Detta resultat erhöles vid hågkomsttestet som genomfördes direkt efter inlärningsfasen av swahili-svenska ordpar och replikerades sedan en vecka efter, och fyra veckor efter initial inläring. Därtill observerades en signifikant testeffekt både avseende ordpar som deltagarna inte hade exponerats för sedan inläringstillfället (de unika ordparen), och ordpar som delvis hade förekommit i de tidigare hågkomsttesten (de ackumulerade ordparen). Detta stämmer överens med resultatet som observerades i experiment 2 av Bertilsson och kollegor (2017) som använde en likartad studiedesign men som hade ett lägre antal deltagare ($n = 29$). Resultatet i denna studie bekräftar att testeffekten håller i sig utan att man exponeras för materialet (unika ordpar) så långt som fyra veckor efter inläringen. Det bör dock framhållas att det låga antalet korrekt återgivna ordpar vid hågkomsttestet på de unika ordparen fyra veckor efter inläringstillfället medför viss minskning av den statistiska kraften, samt en begränsning i statistiska analyser. Framtida studier som tar detta i beaktande behövs för att få ytterligare klarhet i frågan.

Ett annat syfte för den aktuella studien var att undersöka huruvida individuella variationer avseende AMK, NFC och grit predicerar testeffektens magnitud. Regressionsanalysen visade att AMK inte hade någon signifikant inverkan på testeffektens magnitud vid något av hågkomsttesten, för varken de unika eller ackumulerade ordparen. Denna observation motsäger Agarwal och kollegor (2016) som fann att lägre AMK var associerat med större fördel av testbaserad inläring. Resultatet visar å andra sidan inte heller att 'rika blir rikare' (Rapport, Brines, Axelrod, & Theisen, 1997), som i det här fallet skulle innebära att deltagare med högre AMK har mer inneboende resurser att utnyttja för att dra nytta av testbaserad inläring, jämfört med deltagare med lägre AMK (Brewer & Unsworth, 2012). Istället indikerar resultatet att testbaserad inläring är likvärdigt fördelaktig för elever oberoende av individuella variationer avseende AMK, vilket går i linje med ett flertal andra studier (Brewer och Unsworth, 2012; Wiklund-Hörnqvist m. fl., 2014; Tse & Pu, 2012).

Vidare visade regressionsanalysen att individuella variationer avseende NFC inte heller hade någon signifikant inverkan på testeffektens magnitud. Detta går emot hypotesen om att testbaserad inläring bör ge större fördelar till personer med högre NFC. Argumentationen till hypotesen var att testbaserad inläring rimligen kräver mer tankeverksamhet än läsning, och att skillnaden i engagemang mellan en tankekrävande uppgift och en uppgift som inte kräver särskilt mycket tänkande är större ju högre NFC en person har. Det är möjligt att skillnaden i hur mycket tankeverksamhet respektive inlärningsmetod kräver inte är tillräckligt stor för att detta samband ska gälla. En annan

potentiell förklaring är att inläring av swahiliord, oavsett betingelse, inte kräver särskilt mycket tankeverksamhet eller problemlösningsförmåga. Detta skulle kunna innebära att betydelsen av NFC inte aktualiseras i den föreliggande studien på grund av materialets natur. Med ett material som kräver mer tankeverksamhet eller problemlösningsförmåga är det möjligt att deltagare med högre NFC blir mer engagerade i testbetingelsen än i läsbetingelsen, och att deltagare med lägre NFC inte skulle skilja sig väsentligt avseende engagemang mellan de två betingelserna. Mer forskning behövs för att kunna fastställa hur relationen mellan NFC och testeffekten ser ut.

Inte heller grit uppvisade någon signifikant inverkan på testeffektens magnitud. Detta skulle kunna bero på att grit helt enkelt inte är relevant för de inlärningsmekanismer som reglerar testeffekten. Men det skulle också kunna bero på det resonemang som motiverade hypotesen, det vill säga att experimentet genomfördes under för kort tid för att skillnader avseende relation till långsiktiga mål ska ha någon mätbar effekt. Det kan därtill ifrågasättas om studiens metod är lämplig för att undersöka betydelsen av grit ur ytterligare en synvinkel. Deltagarna fick ingen prestationsbaserad belöning, och det kan tänkas att de inte ansåg sig ha någon användning av att lära sig några ord på swahili. Detta medför att deltagarna möjligen inte upplevde något större värde av att lyckas bra i uppgiften. Därmed är det befogat att fråga sig huruvida det är rimligt att anta att deltagarna överhuvudtaget satte upp egna lärandemål för uppgiften. Eftersom grit handlar om ihärdighet och passion för långsiktiga mål (Duckworth m. fl., 2007), kan man vidare fråga sig om grit hade haft en annorlunda inverkan ifall deltagarna hade haft större anledning att sätta upp egna lärandemål, och ifall dessa mål hade varit långsiktiga. Framtida studier bör ta hänsyn till detta och överväga att använda en uppgift som sträcker sig över en längre tidsperiod, samt att ge deltagarna incitament med syftet att uppmuntra dem att sätta upp ett långsiktigt lärandemål för sitt utförande av uppgiften. Ett förslag är att använda ett inlärningsmaterial som ingår i deltagarnas utbildning.

Sammanfattningsvis har resultatet i den aktuella studien gett stöd för att testbaserat lärande resulterar i bättre hågkomst än läsning både direkt efter inläring, en vecka efter och fyra veckor efter. Därtill indikerar resultatet att varken individuella variationer avseende AMK, NFC eller grit påverkar hur mycket bättre hågkomst en individ har när de har lärt sig med hjälp av testbaserad inläring jämfört med när de har lärt sig genom läsning. Det behövs dock fler studier för att vetenskapen ska kunna uttala sig med övertygelse om hur individuella skillnader påverkar effekten av testbaserat lärande. För att få vidare insikt i de specifika individuella skillnader som den föreliggande studien undersöker bör framtida studier visa hänsyn till de material- och metodproblem som här har diskuterats.

Testbaserad inläring är överlägsen många andra tekniker för hågkomst (Carrier & Pashler, 1992; Karpicke & Roediger, 2008; Bertilsson m. fl., 2017; Dunlosky m. fl., 2013; Roediger & Karpicke, 2006a). Testeffekten har visat sig vara robust över olika testformat, typer av material, åldrar, tidsintervall, samt i flera representativa utbildningskontexter, och bedöms vara enkel att implementera i klassrummet (Dunlosky m. fl., 2013). Den föreliggande studien visar att testeffekten förefaller vara likvärdigt fördelaktig för personer som skiljer sig åt avseende AMK, NFC och grit, som har visats påverka inläring i flera andra sammanhang (Swanson & Berninger, 1996; Smedt m. fl., 2009; Juffs & Harrington, 2011; Dornic,

Ekehammar, & Laaksonen, 1991; Meier, Vogl, & Preckel, 2014; Stenlund, Jönsson, & Jonsson 2017; Wolters & Hussain, 2015). Testeffektens effektivitet, robusthet, implementerbarhet och jämlikhet över individuella variationer talar för dess lämplighet som undervisningsmetod i klassrummet, och verksamma aktörer inom utbildningsväsendet bör överväga att implementera testbaserad inlärninng i sin verksamhet i syfte av att stärka lärandet.

Referenser

- Agarwal, P. K., Finley, J. R., Rose, N. S., & Roediger III, H. (2016). Benefits from retrieval practice are greater for students with lower working memory capacity. *Memory*, 25(6), 764-771. <https://doi.org/10.1080/09658211.2016.1220579>
- Aickin, M., & Gensler, H. (1996). Adjusting for multiple testing when reporting research results: The Bonferroni vs Holm methods. *American Journal Public Health Association*, 86 (5): 726-728. DOI: 10.2105/AJPH.86.5.726
- Alloway, T.P., & Alloway, R.G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology*, 106, 20–29. doi:10.1016/j.jecp.2009.11.003
- Baddeley, A. (1992). Working memory. *Science*, 255(5944), 556-559. doi: 10.1126/science.1736359
- Bergman-Nutley, S., & Klingberg, T. (2014). Effect of working memory training on working memory, arithmetic and following instructions. *Psychological Research*, 78, 869–877. doi:10.1007/s00426-014-0614-0
- Bertilsson, F., Wiklund-Hörnqvist, C., Stenlund, T., & Jonsson, B. (2017). The testing effect and its relation to working memory capacity and personality characteristics. *Journal of Cognitive Education and Psychology*, 16(3), 241-259. doi:10.1891/1945-8959.16.3.241
- Brewer, G. A., & Unsworth, N. (2012). Individual differences in the effects of retrieval from long-term memory. *Journal of Memory and Language*, 66(3), 407–415.
- Butler, A. C., Karpicke, J. D., & H. L. Roediger (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(4), 918-928. <http://dx.doi.org/10.1037/0278-7393.34.4.918>
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42(1), 116–131. Retrieved from <http://psycnet.apa.org/fulltext/1982-22487-001.pdf>

Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology, 19*, 619–636. doi:10.1002/acp.1101

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition, 34*, 268–276. doi:10.3758/BF03193405

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20*(6), 633–642. <https://doi.org/10.3758/BF03202713>

Congleton, A., & Rajaram, S. (2012). The origin of the interaction between learning method and delay in the testing effect: The roles of processing and conceptual retrieval organization. *Memory & Cognition, 40*, 528–539. doi:10.3758/s13421-011-0168-y

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*(1), 16–29. doi:10.1037/1082-989X.1.1.16

Dornic, S., Ekehammar, B., & Laaksonen, T. (1991). Tolerance for mental effort: Self-ratings related to perception, performance and personality. *Personality and Individual Differences, 12*(3), 313–319. [https://doi.org/10.1016/0191-8869\(91\)90118-U](https://doi.org/10.1016/0191-8869(91)90118-U)

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92*(6), 1087–1101. <http://dx.doi.org/10.1037/0022-3514.92.6.1087>

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (Grit-S). *Journal of Personality Assessment, 91*(2), 166–174. doi:10.1080/00223890802634290

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4–58. doi:10.1177/1529100612453266

Field, A. (2009). *Discovering Statistics Using SPSS*. London, England: Sage Publications Ltd.

Finney, S. J., & DiStefano, C. (2006). Non-normal and categorical data in structural equation. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course*. Greenwich, CT: Information Age.

- Holm, S. (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6 (2), 65-70. doi: 10.4236/ojs.2016.64055
- Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Belmont, CA: Duxbury.
- Huynh, H., & Feldt, L. S. (1976). Estimation of the box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics*, 1, 69-82. doi: 10.2307/1164736
- Juffs, A., & Harrington, M. A. (2011). Aspects of working memory in L2 learning. *Language Teaching*, 44(2), 137-166. <https://doi.org/10.1017/S0261444810000509>
- Karlsson Wirebring, L., Wiklund-Hörnqvist, C., Eriksson, J., Andersson, M., Jonsson, B., & Nyberg, L. (2015). Lesser neural pattern similarity across repeated tests is associated with better long-term memory retention. *Journal of Neuroscience*, 35(26), 9595–9602. doi: 10.1523/JNEUROSCI.3550-14.2015
- Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966–968. doi: 10.1126/science.1152408
- Karpicke, J. D., & Zaromb, F. M. (2010). Retrieval mode distinguished the testing effect from the generation effect. *Journal of Memory and Language*, 62, 227–239. doi:10.1016/j.jml.2009.11.010
- Kuo, T., & Hirshman, E. (1996). Investigations of the testing effect. *American Journal of Psychology*, 109, 451–464. doi:10.2307/1423016
- Meier, E., Vogl, K., & Preckel, F. (2014). Motivational characteristics of students in gifted classes: The pivotal role of need for cognition. *Learning and Individual Differences*, 33, 39–46. doi: 10.1016/j.lindif.2014.04.006
- Nelson, T. O., & Dunlosky, J. (1994). Norms of paired-associate recall during multitrial learning of Swahili-English translation equivalents. *Memory*, 2, 325–335. doi: 10.1080/09658219408258951
- Rapport, L. J., Brines, D.B., Axelrod, B., N., & Theisen, M. E. (1997). Full Scale IQ as mediator of practice effects: The rich get richer. *Clinical Neuropsychologist*, 11(4), 375-380. <http://dx.doi.org/10.1080/13854049708400466>

Roediger, H. L., & Karpicke, J. D. (2006a). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255. doi: 10.1111/j.1467-9280.2006.01693.x

Roediger, H. L., & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*(3), 181–210. <https://doi.org/10.1111/j.1745-6916.2006.00012.x>

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432-1463. <http://dx.doi.org/10.1037/a0037559>

Rowland, C. A., & DeLosh, E. L. (2014). Mnemonic benefits of retrieval practice at short retention intervals. *Memory*. Advance online publication. doi:10.1080/09658211.2014.889710

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. doi: 10.3102/0034654307313795

Smedt, B. D., Janssen, R., Bouwens, K., Verschaffel, L., Boets, B., & Ghesquière, P. (2009). Working memory and individual differences in mathematics achievement: A longitudinal study from first grade to second grade. *Journal of Experimental Child Psychology*, *103*(2), 186-201. <https://doi.org/10.1016/j.jecp.2009.01.004>

Stenlund, T., & Jonsson, B. (2017). Assessing the willingness to elaborate among young students: Psychometric evaluation of a Swedish need for cognition scale. *Frontiers in Education 2*. <https://doi.org/10.3389/feduc.2017.00002>

Stenlund, T., Jönsson, F. U., & Jonsson, B. (2017). Group discussions and test-enhanced learning: Individual learning outcomes and personality characteristics. *Educational Psychology*, *37*(2), 145–156. <http://dx.doi.org/10.1080/01443410.2016.1143087>

Swanson., H. L., & Berninger, V. W. (1996). Individual Differences in Children's Working Memory and Writing Skill. *Journal of Experimental Child Psychology* *62*(2), 358-385. <https://doi.org/10.1006/jecp.1996.0054>

Toppino, T. C., & Cohen, M. S. (2009). The testing effect and the retention interval: Questions and answers. *Experimental Psychology*, *56*, 252–257. doi:10.1027/1618-3169.56.4.252

Tse, C. S., & Pu, X. (2012). The effectiveness of test-enhanced learning depends on trait test anxiety and working-memory capacity. *Journal of Experimental Psychology: Applied*, *18*(3), 253-264. <http://dx.doi.org/10.1037/a0029190>

Unsworth, N., Brewer, G.A., Spillers, G.J. (2013). Working memory capacity and retrieval from long-term memory: the role of controlled search. *Memory and Cognition*, *41*, 242-254. doi: 10.3758/s13421-012-0261-x

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. doi: 10.3758/BF03192720

Wheeler, M. A., Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study versus test trials. *Memory*, *11*, 571–580. doi: 10.1080/09658210244000414

Wiklund-Hörnqvist, C., Jonsson, B., Korhonen, J., Eklöf, H., and Nyroos, M. (2016). Untangling the contribution of the subcomponents of working memory to mathematical proficiency as measured by the national tests: a study among Swedish third graders. *Frontiers in Psychology*, *7*:1062. doi: 10.3389/fpsyg.2016.01062

Wiklund-Hörnqvist, C., Jonsson, B., & Nyberg, L. (2014). Strengthening concept learning by repeated testing. *Scandinavian Journal of Psychology*, *55*(1), 10–16. doi: 10.1111/sjop.12093

Wolters, C.A., & Hussain, M. (2014). Investigating grit and its relations with college students' self-regulated learning and academic achievement. *Metacognition Learning*, *10*, 293-311. 10.1007/s11409-014-9128-9