



Linköpings universitet  
Grundlärarprogrammet, inriktning år F-3

Elin Alderfalk och Vanessa Forsyth

## **Bedömning i matematik i lågstadiet**

En litteraturstudie om bedömningsformer i matematik och reliabiliteten i prov med flervalsfrågor.


Examensarbete 1, inom Ämnesdidaktik  
Matematik  
Forskningskonsumtion

Handledare:  
Mats Bevemyr

LIU-LÄR-G-MA-15/16-SE

---

Institutionen för beteendevetenskap och lärande - IBL

	Institutionen för beteendevetenskap och lärande 581 83 LINKÖPING	<b>Seminariedatum</b>  2015-03-25
--	---	---

<b>Språk</b> Svenska/Swedish Engelska/English	<b>Rapporttyp</b> Examensarbete grundnivå	<b>ISRN-nummer</b> LIU-LÄR-G-MA-15/16-SE
---	--	---

<p><b>Titel</b>          Bedömning i matematik i lågstadiet          En litteraturstudie om bedömningsformer som främjar elevers utveckling i matematik och reliabiliteten i ett prov med flervalsfrågor.</p> <p><b>Title</b>          Assessment in primary school mathematics – A literature review of assessment methods and reliability in multiple-choice tests.</p> <p><b>Författare</b>          Elin Alderfalk och Vanessa Forsyth</p>
--

<p><b>Sammanfattning</b></p> <p>Syftet med denna litteraturstudie är att ta reda på vilken bedömningsform som främjar elevers utveckling i matematik samt hur reliabiliteten i ett matematikprov påverkas av flervalsfrågor. Bedömning har lika stor del i undervisningen som kunskapsförmedlingen. Läraren kan planera sin undervisning utifrån bedömningarna som görs och därmed lägga undervisningen på en väl anpassad nivå för sina elever. Bedömning kan ske formellt som exempelvis vid prov där olika frågeformuleringar påverkar reliabiliteten i proven. Detta är en forskningskonsumtion som grundas på en systematisk litteraturstudie. Metoden vi använt oss av är databassökning i databaserna ERIC och UniSearch. Resultatet visar att formativ bedömning är den mest givande bedömningsmetoden för eleverna och genom kontinuerlig återkoppling främjas elevernas lärande. Den formativa bedömningsprocessen kräver dock mycket tid och engagemang från läraren vilket gör att arbetsbördan och stressnivån ökar. Resultatet visar även att prov med flervalsfrågor har en låg reliabilitet jämfört med andra provtyper. Det går dock att vidta åtgärder som höjer reliabiliteten.</p>
--

<p><b>Nyckelord</b>          Matematik, lågstadiet, bedömning, formativ, flervalsfrågor, reliabilitet</p>
---

# Innehållsförteckning

1.0 Inledning .....	5
2.0 Syfte och frågeställningar .....	6
3.0 Bakgrund .....	7
3.1 Vad är matematik? .....	7
3.2 Bedömning .....	8
3.2.1 Formativ och summativ bedömning.....	9
3.2.2 Formell och informell bedömning.....	9
3.2.3 Återkoppling.....	10
3.3 PISA och TIMSS .....	10
3.4 Frågor med egna konstruerade svar och flervalsfrågor .....	11
3.5 Relevanta begrepp inom bedömning .....	12
3.5.1 Reliabilitet .....	12
3.5.2 Validitet .....	13
4.0 Metod.....	14
4.1 Litteraturstudie och litteratursökningar .....	14
4.2 Källkritik .....	16
4.3 Urval.....	16
4.4 Metoddiskussion.....	19
5.0 Resultat .....	21
5.1 Bedömningsformer som främjar elevernas utveckling i matematik.....	21
5.1.1 Lärares val av bedömningsformer .....	24
5.2 Summering av resultat om bedömningsformer .....	25
5.3 Flervalsfrågors betydelse i matematikprov.....	26
5.4 Summering av resultat om skriftliga frågor.....	30
6.0 Diskussion .....	32
6.1 Formativ bedömning påverkar elevers motivation.....	32
6.2 Bedömningsformer och läraren .....	32

6.3 Återkoppling är bra för eleverna .....	34
6.4 Flervalsfrågor i prov i lågstadiet.....	35
7.0 Slutsats .....	38
8.0 Förslag på vidare forskning .....	38
9.0 Referenslista .....	39

## 1.0 Inledning

Vi är två lärarstudenter på Linköpings Universitet som studerar till grundlärare i F-3. Tillsammans har vi valt att göra en studie (se bilaga 1) som handlar om bedömningsformer och hur flervalsfrågor påverkar tillförlitligheten i prov.

I vår verksamhetsförlagda utbildning har vi sett bristande faktorer som till exempel att läraren har haft dåliga förutsättningar för att ge en god bedömning i matematikundervisningen. Den största delen av bedömning sker via rättning av matteböcker. På det sättet får inte eleverna återkoppling som främjar deras matematikutveckling utan oftast bara en symbol eller kort kommentar i matteboken. Det blir även svårt för läraren att ge en främjande återkoppling när han/hon använder prov med flervalsfrågor. Detta beror på att eleverna inte formulerar sina egna svar utan enbart sätter ett kryss i en ruta. Läraren ser då inte hur eleverna har kommit fram till sina svar och kan därmed endast ge en återkoppling om antal rätt och fel. Flervalsfrågor bidrar därmed till att läraren inte kan ge en god bedömning som främjar elevernas matematikutveckling.

En annan anledning till att läraren inte kan ge en god bedömning utifrån ett prov med flervalsfrågor är att eleverna har möjlighet att gissa. Detta leder till att när läraren rättar provet kan eleverna få rätt på en fråga trots att de egentligen inte har kunskaperna. När läraren gör bedömningen kan han/hon få fel uppfattning av elevernas kompetens.

Två undersökningar som använder flervalsfrågor för att mäta elevers kunskaper i matematik är PISA och TIMSS. Undersökningarna visar att svenska elever har presterat sämre i matematik de senaste åren (Skolverket, 2012). Därför undrar vi om resultaten från undersökningarna kan ha påverkats av flervalsfrågorna. Det PISA och TIMSS kommit fram till har uppmärksammats i samhället och i skolpolitiken vilket gör det till en aktuell fråga i skolan. Trots olika satsningar som exempelvis mattelyftet visar resultaten i undersökningarna inte någon förbättring vilket är anmärkningsvärt. Mycket fokus verkar läggas på undervisningen och därför upplever vi att bedömningsdelen hamnar i bakgrunden. Det här kan verka konstigt eftersom undervisning och bedömning går hand i hand. I denna studie kommer vi därför forska om hur tillförlitligheten i matematikprov påverkas av flervalsfrågor och hur läraren kan ge en främjande bedömning.

## 2.0 Syfte och frågeställningar

Syftet med denna litteraturstudie är att uppmärksamma nuvarande och blivande lärare hur användning av olika bedömningsformer påverkar elevers utveckling i matematik samt hur reliabiliteten påverkas i prov av flervalsfrågor. Med denna studie vill vi bidra med en djupare kunskap om hur lärarens val av bedömningsform kan främja elevers utveckling i grundskolans tidigare år. Vi vill även lyfta fram hur läraren kan göra en bedömning för att öka reliabiliteten i prov med flervalsfrågor.

Frågeställningar:

- Vilka bedömningsformer främjar elevernas fortsatta utveckling i matematik?
- Hur påverkas reliabiliteten i ett matematikprov av flervalsfrågor?

## 3.0 Bakgrund

I detta avsnitt börjar vi med att beskriva vad matematik är. Vi går sedan vidare med att redogöra för olika bedömningsformer. Eftersom PISA och TIMSS innehåller de typer av frågor vi ska forska om berättar vi lite kort om dessa undersökningar. I samband med detta fördjupar vi oss i de olika typer av frågor som jämförs i studien. Slutligen kommer vi att definiera viktiga begrepp som studien kommer att innehålla.

### 3.1 Vad är matematik?

Enligt Roos och Kiselman (2015) är matematik vetenskapen för att lösa problem kring beräkningar av olika slag och utveckla metoder för detta, varvid man använder olika så kallade matematiska teorier. Matematik handlar om tal och deras egenskaper samt förhållandet mellan dem. Roos och Kiselman (2015) beskriver författarna också matematik som ett språk som människan använder för att tala om kvantiteter, dvs. mängder som kan räknas eller mätas som därför uttrycks som tal. De nämner också hur matematik är en abstrakt och generell vetenskap för problemlösning och metodutveckling.

I den akademiska världen passar det bra att definiera matematik som en abstrakt vetenskap men inom grundskolans matematik kan denna definition ge problem. I grundskolan är det viktigt att matematiken är konkret med modeller som eleverna kan falla tillbaka till (Löwing & Kilborn, 2002).

Enligt läroplanen (Skolverket, 2011a) är matematik en kreativ, reflekterande och problemlösande aktivitet som utvecklas genom människans nyfikenhet och lust att utforska samt för att uppfylla människans praktiska behov. Människan behöver kunskaper i matematik för att kunna fatta beslut i vardagslivet och ökar därmed möjligheterna att aktivt delta i samhällets beslutsprocesser.

I rapporten *Hög tid för matematik* nämner Johansson (2001) att det finns sju olika förmågor som innebär ett rikt och nyanserat kunnande i matematik. Målet i all undervisning är att utveckla intresse, nyfikenhet och förmågor hos alla elever så att de därmed lär sig den grundläggande matematiken. De sju förmågorna som Johansson (2001) nämner för ett mångfasetterat kunnande är:

- *Produktivt förhållningssätt*  
 Detta innebär att eleven ser matematiken som någonting meningsfullt och användbart i vardagslivet och i samhället.
- *Helhetsperspektiv*  
 Med ett helhetsperspektiv har eleven förmågan att värdera matematikens roll i ett historiskt, kulturellt och samhälleligt perspektiv.
- *Begreppslig förståelse*  
 Den begreppsliga förståelsen innebär att eleven förstår innebörden av de begrepp som ingår i matematik.
- *Behärskande av procedurer*  
 Förmågan att behärska procedurer innebär att eleven kan använda sig av olika procedurer och algoritmer för att komma fram till lösningar.
- *Kommunikationsförmåga*  
 Eleven kan föra diskussioner i både tal och skrift kring frågeställningar i matematik.
- *Strategisk kompetens*  
 Med en strategisk kompetens kan eleven formulera, representera och lösa vardagliga och relevanta matematiska problem.
- *Argumentationsförmåga*  
 Eleven kan tänka logiskt, reflektera och förklara hur han/hon har tänkt samt argumentera för sin sak.

För att lärare ska kunna ge rättvisa bedömningar och ha kännedom om varje elevs kunskapsutveckling måste läraren göra både informella och formella bedömningar. Det är genom bedömningarna som läraren kan avgöra om eleverna har dessa sju förmågor eller inte.

### 3.2 Bedömning

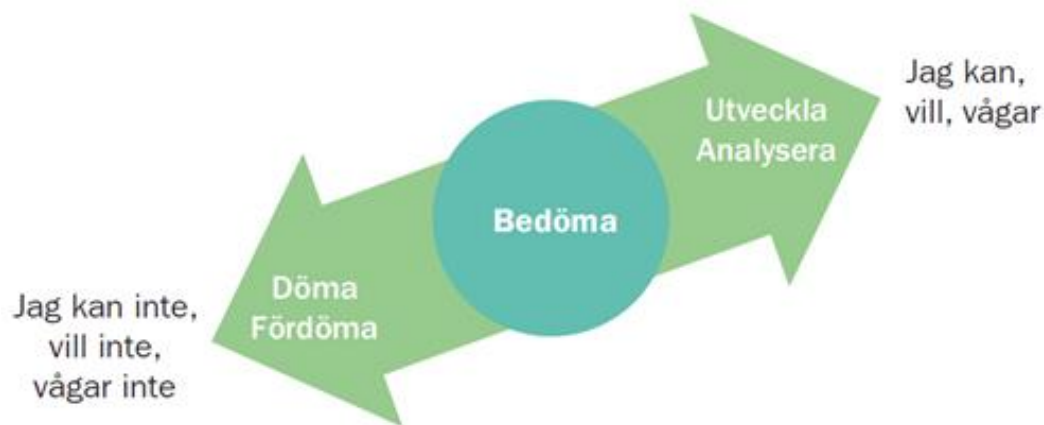
Bedömning kan ses som den process när läraren samlar in information och utvärderar elevernas kunskaper (Skolverket, 2011b). Det finns flera olika bedömningsformer som läraren kan använda sig av i skolan för att få syn på elevernas utveckling. Provet kan se väldigt olika ut där det t.ex. kan finnas korta frågor att svara på eller öppna frågor där eleven kan föra en diskussion om sin nyvunna kunskap (Skolverket, 2011c).



### 3.2.1 Formativ och summativ bedömning

Formativ bedömning är en typ av bedömning som syftar till att stärka elevens lärande medan en summativ bedömning handlar om att summera elevens kunskaper i ett omdöme eller betyg (Woolfolk, 2014). Den formativa bedömningen karakteriseras av att undervisningens mål klargörs för eleven, att läraren vet var eleven befinner sig i förhållande till målet och att eleven får återkoppling som syftar till att ge eleverna vägledning om hur målet ska uppnås (Skolverket, 2010). Den formativa bedömningen är en del av undervisningen och kan även ske genom självbedömning eller kamratbedömning. Självbedömningen handlar om att eleven själv kan se sina styrkor och svagheter och reflektera över sitt eget arbete. På detta sätt kan eleven själv se nästa steg i sin utveckling och vet då hur han/hon ska gå vidare i arbetet. Kamratbedömning kan också fungera som en typ av formativ bedömning där en klasskompis utvärderar ett arbete och ger framåtsyftande återkoppling till den elev som gjort arbetet (Woolfolk, McMillan, Hughes, & Walkup, 2011).

Denna figur som är hämtad från Skolverket (2014b) visar att eleven kan påverkas både positivt och negativt beroende på hur läraren ger bedömning. Bedömningskulturen i skolan spelar alltså stor roll för elevens inställning till fortsatt lärande.



Figur 1. Bedömningens påverkan

### 3.2.2 Formell och informell bedömning

Bedömningar av elevernas lärande kan ha olika karaktär, de kan nämligen vara formella eller informella. Formella bedömningar innebär att eleven vet om att han/hon blir bedömd vid ett visst tillfälle, exempelvis vid ett prov eller en redovisning. Informella bedömningar sker istället i samband med undervisningen när läraren observerar och diskuterar med eleven för att få en uppfattning om hur eleven ligger till (Woolfolk m.fl., 2011).

### 3.2.3 Återkoppling

Återkoppling är en viktig del i arbetet med att bedöma eleverna. När läraren ger eleverna kontinuerlig återkoppling får eleverna en tydligare bild av hur de ligger till och vad de behöver göra för att förbättra sig. Återkoppling gör också att läraren tidigt ser om någon elev inte förstår och halkar efter och kan därmed förändra sin undervisning för att få med sig alla elever. Det som är viktigt för läraren att tänka på vid återkoppling till eleverna är att återkopplingen ska innehålla både feedback och feed-forward. Feedback är den återkoppling som syftar till att berätta vart eleven befinner sig just nu och feed-forward är den återkoppling som syftar till hur eleven ska ta sig vidare. I återkopplingen finns det därmed både en summativ och en formativ del (Skolverket, 2011d). Det finns även en tredje form av återkoppling som kallas för feed-up. Denna typ av återkoppling är kopplad till vilka mål eller kunskapskrav som eleverna ska uppnå (Björklund Boistrup, 2010).

### 3.3 PISA och TIMSS

PISA står för *programme for international student assessment* och är en internationell studie. Denna studie undersöker hur utbildningssystemet bidrar till att femtonåriga elever utvecklar kunskaper inom matematik, naturvetenskap och läsförståelse som dem behöver senare i livet. Studien ger ett underlag för alla deltagande länder att uppmärksamma sina egna styrkor och svagheter inom utbildningen. Denna observation kan sedan vägleda länderna till en förbättring av skolan. Studien sker vart tredje år (Skolverket, 2014a). En annan internationell studie är TIMSS som står för *Trends in International Mathematics and Science Study*. Denna studie syftar till att undersöka elevers kunskaper i matematik och naturvetenskap i både årskurs fyra och åtta vart fjärde år (Skolverket, 2012).

Resultaten från den senaste TIMSS undersökningen 2011 visar att eleverna i årskurs fyra ligger under genomsnittet i OECD-länderna. Resultaten visar även att eleverna i årskurs åtta ligger under genomsnittet och denna försämring har pågått ända sedan år 1995. Elever som deltog i TIMSS undersökningen när de gick i fjärde klass år 2007 deltog även när de gick i årskurs åtta 2011. Eftersom samma elever deltog vid båda tillfällena är det möjligt att jämföra kunskapsutvecklingen hos elever i Sverige och elever i andra länder. Jämförelsen visar att elever i andra länder har lärt sig mer än elever i Sverige under de fyra åren (Skolverket, 2012).

Resultaten har sedan första PISA-undersökningen, år 2000, visat en nedåtgående trend i Sverige. Den senaste undersökningen år 2012 visade att svenska elever presterade under

OECD-genomsnittet i alla tre kunskapsområden. I PISA 2012 presterar 25 av 34 länder bättre i matematik än Sverige. Det är inte bara de lågpresterande eleverna som tappar i matematikkunnande utan även bland de högpresterande eleverna är nedgången stor (Skolverket, 2014c).

I både TIMSS och PISA undersökningarna förekommer två olika slags frågeformuleringar som eleverna får besvara. Det är flervalsfrågor och frågor där eleverna får konstruera svaren själva (Skolverket, 2014d).

### 3.4 Frågor med egna konstruerade svar och flervalsfrågor

I den här studien jämförs de två olika frågeformuleringarna som förekommer i PISA och TIMSS undersökningarna, vilka är flervalsfrågor och frågor där eleven konstruerar sitt eget svar. Det som skiljer dessa åt är att en flervalsfråga har givna svarsalternativ som eleven får välja mellan medan en fråga där eleven konstruerar sitt eget svar, d.v.s. en öppen/sluten fråga, är en fråga där eleven får svara fritt. I matematikböcker i lågstadiet förekommer oftast frågor där eleven får svara själv, de kan ofta se ut såhär:

$$”23 + 57 = \underline{\hspace{2cm}}”$$

Denna formulering kan också refereras till en sluten fråga, då frågan har endast ett rätt svar. Svaret på den här slutna frågan kan antingen vara rätt, det vill säga 80, eller fel vilket kan vara vilket annat tal som helst. Det förekommer också frågor av denna typ där eleven får skriva lite längre, eller visa sina beräkningar. Dessa kan ses mer som öppna frågor där svaren kan besvaras på lite olika sätt. Fördelen med frågor där eleven konstruerar sitt eget svar är att eleven visar konkret hur den har tänkt och på det sättet, även om svaret är felaktigt, kan läraren få en inblick i hur eleven ändå har tänkt. En nackdel med detta provformat är att det tar längre tid för en lärare att rätta proven då han eller hon måste hantera varje svar olika vilket kräver mer arbetstid, tid som egentligen skulle kunna gå åt planering för nästa lektion eller dylikt (Ventouras, Triantis, Tsiakas & Stergiopoulos, 2009).

Det som definierar en flervalsfråga är att det finns en fråga med två eller flera svarsalternativ att välja mellan. Oftast är ett av alternativen rätt men vid vissa tillfällen kan det finnas flera alternativ som kan ses som korrekta. Ett exempel på en flervalsfråga kan vara:

*”Lisa tar bussen 2,5 km in till staden vilket tar 15 minuter. Sedan tar Lisa bussen vidare till landet som ligger 5km bort. Hur lång tid tar det för Lisa att åka till landet?*

*A. 15 minuter*

*B. 25 minuter*

*C. 35 minuter*

*D. 45 minuter”*

Även vid flervalsfrågor finns det för- och nackdelar. Fördelen är att det går betydligt mycket snabbare att rätta ett prov med flervalsfrågor. Nackdelen är att läraren bedömer i det fallet endast elevens kompetens i att ge rätt svar och inte hur den kom fram till svaret (Ventouras m.fl., 2009). Genom att eleven exponeras för olika svarsalternativ får eleven en procentuell sannolikhet att svara rätt på en fråga även fast den inte kan svaret. I Sverige förekommer ingen rättningsmetod som gör att lärare kan utesluta flervalsfrågor som besvarats genom gissning. I USA bl.a. har de en rättningsmetod som heter ”negative marking” som innebär att eleven får poäng för de korrekt besvarade frågorna och minuspoäng för de inkorrekt besvarade frågorna (Burton, 2004).

### 3.5 Relevanta begrepp inom bedömning:

Det vanligaste problemet som förekommer vid bedömning av prov är att läraren misstolkar resultaten (Woolfolk, 2014). Det uppträder när läraren tror att provet noggrant mäter elevens kunskaper. Woolfolk (2014) nämner att inget prov kan mäta elevens exakta kunskaper. Syftet med ett prov kan variera och för att ta hänsyn till att provet ska uppnå sitt syfte krävs reliabilitet och validitet (Woolfolk, 2014).

#### 3.5.2 Reliabilitet

Reliabiliteten definierar bedömningsresultatens tillförlitlighet och stabilitet. Det finns olika metoder för att förbättra reliabiliteten, som till exempel att ha flera frågor som mäter samma sak eller flera bedömare. Är det hög reliabilitet ska resultaten bli detsamma vid upprepade mätningar (Woolfolk, 2014). Är det låg reliabilitet innebär det att mätproceduren lätt kan påverkas av metodfel, som till exempel genom oklarheter i frågeformulering eller bristande instruktioner om hur provfrågorna ska besvaras. Ett mätinstrument för detta är till exempel test-retest-metoden vilket innebär att man utför flera mätningar med samma test på samma individer (Woolfolk, 2014). Om en lärare har ett matematikprov bör läraren ha flera prov som

mäter samma sak eller åtminstone flera frågor som mäter samma sak för att få högre reliabilitet på provet.

### 3.5.1 Validitet

Woolfolk (2014) poängterar att ifall ett prov är tillförlitligt är nästa fråga om det har validitet. Validitet är huruvida bedömningen, i det här fallet provet, mäter det som det är avsett att mäta, som till exempel ett matematikprov med skriftliga frågor. En elev kanske har de matematiska kunskaperna men brister i den språkliga delen vilket gör att eleven kan få problem att visa sina matematikkunskaper. Det finns tre olika aspekter som är viktiga att beakta vid värdering av ett provs validitet. Dessa tre aspekter är innehållsvaliditet, kriterievaliditet och begreppsvaliditet. Innehållsvaliditet innebär att provet består av frågor som täcker varje komponent som ska mätas. Kriterievaliditet illustrerar i vilken utsträckning ett prov mäter samma sak som ett annat prov. Begreppsvaliditet innebär att de begrepp som provet avser att mäta provas (Woolfolk, 2014).

## 4.0 Metod

Detta arbete är en forskningskonsumtion vilket innebär att vi bearbetar och analyserar befintlig forskning från nationella och internationella studier. Bearbetningen och analysen sker med utgångspunkt för syftet och frågeställningarna. Vi kommer i denna del att ta upp vad en litteraturstudie är och hur en litteratursökning går till. Vi kommer därefter att gå in på vilken hänsyn vi tagit till källkritik och hur urvalet av våra artiklar har gjorts. Till sist följer en metoddiskussion där vi tar upp faktorer som har påverkat urvalet.

### 4.1 Litteraturstudie och litteratursökningar

En litteraturstudie är en studie där författaren ska sammanställa litteratur kring befintlig forskning samt analysera och diskutera den (Hartman, 2003). I en litteraturstudie ställer författaren en eller flera frågor som den vill ha svar på och det är dessa frågeställningar som ligger till grund för vad som återges av litteraturen (Hartman, 2003).

Enligt Eriksson Barajas, Forsberg och Wengström (2013) finns det olika typer av litteraturstudier som exempelvis, allmän litteraturstudie, systematisk litteraturstudie eller begreppsanalys. Vi har gjort en systematisk litteraturstudie då vi främst har sökt systematiskt efter befintlig forskning för att besvara våra frågeställningar. Det som skiljer en systematisk litteraturstudie från en allmän litteraturstudie är att författaren tydligt redovisar metoderna som har använts och att de är öppna för granskning. Författaren nämner att en systematisk litteraturstudie syftar till att identifiera all befintlig evidens som är väsentlig för ett genomgående ämne. En systematisk litteraturstudie strävar därmed efter att heltäckande identifiera, bedöma och syntetisera alla relevanta studier inom ett givet ämne (Eriksson Barajas m.fl., 2013).

Hartman (2003) lägger vikt vid att litteratursökningen är en viktig del av ett forskningsarbete. Processen utförs genom att författaren formulerar sökord som utgör en grund för litteratursökningen (Eriksson Barajas m.fl., 2013). Sökningen kan sedan gå till på olika sätt genom en manuell sökning eller en databassökning. Manuella sökningar kan utföras på olika sätt. Om till exempel författaren har hittat en artikel som rör problemområdet bör referenslistan studeras i syfte till att finna andra artiklar som berör problemområdet. En annan typ av manuell sökning kan vara att författaren granskar innehållsförteckningen i en tidsskrift som är specialiserad på det valda ämnet, som till exempel litteraturdidaktik eller matematik. Eriksson Barajas m.fl. (2013) nämner ett tredje sätt att söka manuellt efter artiklar och det

innebär att författaren undersöker forskares hemsidor på lärosätena i syfte att leta efter pågående aktuell forskning.

Om författaren gör en databassökning finns det många olika databaser att söka litteratur i, ett exempel är databasen ERIC som är en bred databas som innehåller litteratur om pedagogik och psykologi (Eriksson Barajas m.fl., 2013). Problemet som oftast uppstår vid litteratursökningar är överflödet av material (Hartman, 2003). Det krävs då att författaren använder olika sökstrategier för att enklast avgränsa antal träffar i databassökningen (Eriksson Barajas m.fl., 2013). Eriksson Barajas m.fl. (2013) skriver om konsten att söka fakta i databaser som innebär att man formulerar rätt frågor, avgör hur gamla arbeten som ska ingå, vilket språk de ska vara på, samt vilka studietyper som är intressanta. Kriterier för sökningen bestäms utifrån de frågeställningar författaren har ställt. Enligt Eriksson Barajas m.fl. (2013) är nästa steg att välja databas och sökord som ska användas. Oftast väljs sökbegrepp utifrån ord i frågeställningarna och så söker författaren på enstaka ord eller ordkombinationer av logiska sökoperatorer. Sökoperatorerna kombinerar sökorden i en fritextsökning med hjälp av ”AND”, ”OR” och ”NOT” (Eriksson Barajas m.fl., 2013).

Det finns sex steg i urvalsprocessen enligt Eriksson Barajas m.fl. (2013) som vi följt vid sökningen av artiklarna, dessa är:

- *Steg 1*  
Identifiera intresseområde och definiera sökord.
- *Steg 2*  
Bestäm kriterier (tidsperiod och språk) för vilka studier som ska väljas.
- *Steg 3*  
Genomför sökning i lämpliga databaser.
- *Steg 4*  
Sök på egen hand efter ej publicerade artiklar för att finna pågående forskning inom området.
- *Steg 5*  
Välj relevanta titlar och läs sammanfattningar (abstracts). Gör ett första urval av litteratur som ska bli föremål för fortsatt granskning.
- *Steg 6*  
Läs artiklarna i sin helhet och gör en kvalitetsvärdering.

## 4.2 Källkritik

Källkritik handlar om att kontrollera fakta från en källa. En källa kan vara skriftlig, muntlig eller materiell och är ursprunget till vår kunskap. Allt som vi hämtar vår kunskap ifrån är kunskapskällor och källkritiken syftar till att värdera och bedöma dessa källors trovärdighet. Det handlar om att ta reda på vad som är sant, eller i varje fall sannolikt (Thurén, 2005).

Källkritiken består enligt Thurén (2005) av fyra källkritiska principer för att ta reda på detta:

- *Äkthet*

Handlar om att källan är vad den utger sig för att vara, att den är äkta och därmed inte en förfälskning.

- *Tidssamband*

Handlar om att en källa är mer trovärdig ju mer aktuell den är. Om det gått lång tid mellan en händelse och en källas berättelse om händelsen finns det anledning att tvivla på källan.

- *Oberoende*

Handlar om att källan ska stå för sig själv och inte vara exempelvis en avskrift av en annan källa. Man brukar skilja mellan förstahandskällor och andrahandskällor eller primär- och sekundärkällor. Primärkällan är den ursprungliga källan och är därför mest trovärdig medan sekundärkällan har traderats och därmed gått i flera led. Det finns dock undantag och en aktuell sekundärkälla kan vara mer trovärdig än en gammal primärkälla.

- *Tendensfrihet*

Handlar om att källan är partisk och har ett eget intresse av att ge en felaktig bild av verkligheten. När personen som sprider information själv kan vinna på det finns det skäl att misstänka att personens egna intressen går före sanningen.

Dessa fyra principer har vi haft i åtanke när vi valt ut och bearbetat artiklarna. Detta har gjort att de slutgiltiga artiklarna anses vara trovärdiga och är aktuella.

## 4.3 Urval

I artikelsökningen utgick vi från de sex stegen som Eriksson Barajas m.fl. (2013) tar upp. Efter att vi hade valt intresseområde bestämde vi sökord och kriterier för att hitta lämpliga artiklar. De sökord vi använde oss av är främst ord som finns med i våra frågeställningar som exempelvis *bedömning (assessment)*, *matematik (mathematics)* och *reliabilitet (reliability)*. Detta gjorde vi för att öka chanserna att få artiklar som berör just våra frågeställningar. För att



minska antalet träffar lade vi till ord som hör till frågeställningarna som exempelvis *flervalsfrågor (multiple choice questions)*, *formativ (formative)* och *summativ (summative)*. Vidare valde vi att begränsa vår sökning genom att välja artiklar mellan år 2000 och 2014. Detta gjorde vi för att inte stöta på för gammal forskning. De språk vi kunde tänka oss att läsa var engelska och svenska, artiklarna vi till slut valde att använda var dock endast på engelska. För att enbart få fram vetenskapliga och därmed noggrant granskade artiklar valde vi att kryssa i *peer reviewed*. Databaserna som användes och gav resultat var ERIC (Educational Resources Information Center) och UniSearch. ERIC är en bred databas som täcker forskning inom områdena psykologi och pedagogik (Eriksson Barajas m.fl., 2013). UniSearch är däremot en katalog där träffar från flera olika databaser kommer upp när man söker. Hartman (2003) menar att det är en fördel att söka i dessa typer av kataloger då de är lättillgängliga och därför valde vi att använda denna sökningsmetod.

Ett exempel på hur en av våra sökningar gick till är att vi började med att använda sökordet ”mathematics” i databasen ERIC och fick fram 86,734 träffar. För att minska ner antalet träffar lade vi till sökordet ”multiple choice questions” och fick då 284 träffar. Tredje steget var att vi kryssade i ”peer reviewed” och fick då 78 träffar. Eftersom vi ville ha artiklar mellan åren 2000 till 2014 valde vi det och därefter fick vi 58 träffar. Efter att ha fått 58 träffar gick vi igenom titlarna för att sälla ut ytterligare artiklar. Tillslut hade vi 7 artiklar kvar som vi valde att läsa abstracten på och fick då fram en artikel som var relevant för frågeställningen. Vi upprepade samma process med olika kombinationer av sökord för att få fram fler artiklar till studien. När vi hade valt ut några artiklar som utifrån abstracten verkade relevanta för studien läste vi hela artiklarna för att se vilka vi kunde använda oss av. Tillslut valde vi elva stycken artiklar som vi presenterar i arbetet för att kunna svara på våra frågeställningar på ett så bra sätt som möjligt. För att ge en överblick av vilka artiklar vi hittade har vi samlat dem i tabell 1 nedan.

Tabell 1. Sammanställning av artiklar

Författare	Årtal	Land	Databas	Sökord	Metod	Deltagare
Tveit, S.	2013	Norge	ERIC	Mathematics + formative	Litteraturstudie	
Antoniou, P. James, M.	2014	Cypern	UniSearch	Mathematic + assessment + primary + formative	Intervju, icke-deltagande observation	4 lärare från 2 skolor
Veldhuis, M., van den Heuvel-Panhuizen, M.	2014	Nederländerna	UniSearch	Primary School + Assessment + Mathematics + Education	Enkät	960 lärare från 557 skolor
Aureliana-Loredana, P.	2014	Rumänien	UniSearch	Mathematic + assessment + primary + formative	Litteraturstudie	
Ginsburg, H.	2009	USA	UniSearch	formative + assessment + mathematics + education	Litteraturstudie	
Ventouras, E., Triantis, D., Tsiakas, P., Stergiopoulos, C.	2009	Grekland	ERIC	Comparison + multiple choice	Experiment	63 elever
Hastedt, D., Sibbers, H.	2005	Tyskland	ERIC	Math + test + multiple choice questions	Dataanalys	
Lau, P. N. K., Lau, S. H., Hong, K. S. & Usop, H.	2011	Malaysia	ERIC	Multiple choice + guessing	Experiment	449 elever från 19 skolor
Roediger III, H., Marsh, E.	2005	USA	ERIC	Multiple choice + guessing	Experiment	24 elever
Jonsson, A., Lundahl, C., Holmgren, A.	2014	Sverige	UniSearch	Bedömning + matematik	Experiment	900 lärare från 40 skolor
Burton, R.	2004	Storbritannien	ERIC	Multiple choice + guessing	Analys	

#### 4.4 Metoddiskussion

För att kunna göra en litteraturstudie krävs att man hittar flera relevanta artiklar om forskning som gjorts för att besvara studiens frågeställningar. Vetenskapliga artiklar har ett tydligt upplägg där vissa delar ska finnas med som exempelvis metod och resultat (Eriksson Barajas m.fl., 2013). Vi har i våra sökningar stött på artiklar där inte alla delar finns med och därför har de artiklarna uteslutits. Vi hade också krav på att artiklarna vi hittade skulle vara från tidigast år 2000, vilket innebär att artiklar från tidigare år inte har kommit upp i sökningarna. Vid urvalsprocessen läste vi till en början endast titlarna på artiklarna och om titeln verkade relevant för studien valde vi att läsa artiklarnas abstract. De artiklar som inte hade en tydlig titel valdes därmed bort. Detta kan ha medfört att artiklar uteslutits trots att de kunde ha varit användbara. Vi märkte även att artiklar med relevant titel och abstract kunde visa sig vara oanvändbara för vår studie.

Vi har valt ut elva artiklar som syftar till att svara på våra frågor i studien. När det kommer till frågeställningen om flervalfrågor fann vi inte tillräckligt många artiklar där forskningen handlade om matematik eller lågstadiet. Eftersom vi var under tidspress fick vi därför utöka vår sökning till äldre åldrar och andra skolämnen. Vi tog detta beslut eftersom flervalfrågor är utformade på samma sätt oberoende av ålder och ämne samt att möjligheten för gissning förekommer oavsett hur gamla eleverna är. Det förekommer också forskning från äldre åldrar i frågeställningen om bedömningsmetoder. Anledningen till detta är att vi fick intrycket av att artiklarna handlade om forskning i lågstadiet eftersom det stod det i nyckelorden samt abstracten men efter att ha läst hela artikeln fanns bara resultat från äldre åldrar. Som nämnt hade vi tidspress och eftersom bedömningsmetoder påverkar elever på liknande sätt oberoende av ålder valde vi att behålla artiklarna. Vi tar hänsyn till dessa faktorer när vi presenterar resultaten från artiklarna.

Några av artiklarna som behandlar bedömningsformer är litteraturstudier och därmed sekundärkällor. Anledningen till att vi har valt att använda litteraturstudier i vårt arbete beror på att de sammanställer och analyserar forskning från flera källor. Istället för att använda en artikel som innehåller forskning från en författare om bedömningsformer använder vi en litteraturstudie som behandlar forskning från flera olika författare. De resultat vi utgått från är författarnas slutsats i litteraturstudierna. Vi har även en studie som handlar om lärares val av bedömningsformer. Syftet med den är att komplettera vårt resultat genom att se vilka bedömningsformer de flesta lärare använder.

Vid sökningar måste man beakta många olika faktorer som kan påverka sökningens resultat. Sådana faktorer kan till exempel vara artikelns upplägg eller relevans för studien, hur gammal forskningen är samt tidsbrist. På grund av tidsbrist kan det vara nödvändigt att ändra kriterierna för forskningen som exempelvis ämne eller ålder.

## 5.0 Resultat

I detta avsnitt presenterar vi resultatet av vår litteraturstudie genom att besvara våra frågeställningar utifrån passande rubriker. Först redovisas olika bedömningsformer i matematikundervisning och därefter hur flervalfrågor påverkar reliabiliteten i prov.

### 5.1 Bedömningsformer för att främja elevens utveckling i matematik

Antoniou och James (2014) utförde en studie där fyra lärare från tredje och fjärde klass deltog. Undersökningen gjordes på Cypern och skolorna ligger i de två största skoldistrikten i landet. Målet med studien var att utforska vad läraren gör som kan räknas som formativ bedömning samt utveckla en ram av strategier för analys av formativ bedömning. Författarna utförde intervjuer med lärarna före och efter matematiklektionerna för att samla in data till sin studie. De använde sig även av icke-deltagande observationer där de filmade matematiklektioner för att studera hur det vardagliga arbetet fungerade i klassrummet. För att komplettera datainsamlingen till studien analyserades även elevernas arbeten där läraren kommenterat deras texter.

Antoniou och James (2014) utgår från fem olika strategier som hjälper dem att analysera resultatet. Dessa fem analysprocesser handlar om att klargöra och dela förväntningar, samla in information, tolka informationen, ge feedback och reglera elevernas lärande. Resultatet som författarna kom fram till visar att formativ bedömning är något som lärarna vill använda sig mer av men att det är svårt att uppnå. Antoniou och James (2014) menar att lärarna hade positiva uppfattningar om formativ bedömning och de ansåg att det främjar elevernas lärande. Trots detta visar resultaten stora svagheter i lärarnas formativa bedömningar. Många faktorer påverkade lärarnas bedömningar som exempelvis tidsbrist, svårigheter i att förklara kriterier och vaga kommentarer om elevernas styrkor och svagheter. När en elev fick feedback på ett prov förekom det oftast bara en poängsumma, en symbol eller en kort kommentar som ”bra jobbat!”. De flesta av lärarna ställde även fel typ av frågor enligt Antoniou och James (2014). Frågorna som ställdes ledde nämligen inte till vad eleverna förstått eller hur de skulle utveckla sitt lärande. En annan viktig aspekt var att alla elever fick samma uppgifter trots att de låg på olika nivå, vilket ledde till att vissa elever fick stressa för att hinna klart medan andra elever blev uttråkade av uppgifterna. Tre av lärarna kände sig oroliga över att hinna med allt de skulle enligt läroplanen och därför lades tonvikt på detta istället för på elevernas lärande. En av lärarna utvärderade och ändrade dock sin undervisning efter att ha tolkat elevernas lärande.

Denna lärare var den enda som tog till sig konceptet av formativ bedömning och ändrade sin undervisning under tidens gång.

En studie som utförts av Jonsson, Lundahl och Holmgren (2015) handlar också om bedömning för lärande. I studien deltog 900 lärare från 40 olika skolor i Borås under skolåren 2011 och 2012, resultaten är dock endast från gymnasiet då resultat från lägre åldrar inte ännu analyserats. Studien syftade till att införa bedömning för lärande, vilket innebär att se var eleven befinner sig, vart eleven är på väg i sitt lärande och vad eleven behöver göra för att nå målen. Arbetet med att införa bedömning för lärande skedde utifrån fem olika strategier: förklara förväntningar och bedömningskriterier, designa lärandesituationer som visar elevens förståelse, ge konstruktiv feedback, aktivera eleverna som resurser för varandra i sitt lärande samt aktivera eleverna som ägare av sitt eget lärande. Lärarna fick en utbildning under två dagar som grund i bedömning för lärande där de fem strategierna presenterades. Vidare träffades lärarna i grupper om 8-12, ungefär en gång i månaden för att utvärdera undervisningen. Data till studien samlades in från lärarnas möten där skillnader jämfördes från hur bedömningspraktiken såg ut före till efter bedömning för lärande med de fem strategierna infördes. Jonsson m.fl. (2015) gav även ut två olika enkäter, en för lärarna och en för eleverna, som kompletterade resultaten.

Resultatet som Jonsson m.fl. (2015) kom fram till presenteras utifrån fyra olika punkter. Den första behandlar om det blivit någon ökning av pedagogiska diskussioner i skolan. Nästan alla lärare meddelar att diskussionerna ökat, både av formellt och informellt slag. Författarna menar att lärarna diskuterade mer både på formella möten men även i korridorerna i skolan eller vid fiket. Lärarna pratade mer om elevernas lärande och fick större förtroende för att alla elever kan lära sig. Den andra punkten handlar om ökad öppenhet om vilka förväntningar som finns. Författarna menar att eleverna blev mer medvetna om vilka mål som skulle uppnås och fick ett ökat fokus på kvaliteten i sina arbeten. Detta tros komma från att lärarnas dokumentationer hade blivit tydligare liksom även deras sätt att följa upp och kommunicera elevernas framsteg (Jonsson m.fl. 2015). Den tredje punkten handlar om att bedömning för lärande fick ta en större plats i verksamheten. Jonsson m.fl. (2015) lyfter fram att lärarna gav eleverna återkoppling på ett nytt sätt och att de använde sig av bedömningsmatriser, vilket uppskattades av eleverna. Även självbedömning och kamratbedömning utnyttjades mer av lärarna. Den sista punkten rör sig om hur lärarnas arbetsbörda påverkats. Denna punkt är den enda som gett negativa effekter efter införandet av bedömning för lärande. Författarna menar

att lärarna rapporterade om större arbetsbörda och ökad stressnivå. Detta berodde främst på att eleverna krävde fler tillfällen för att revidera texter och mer återkoppling med målet att nå ett högre betyg. Lärarnas arbetsbörda ökade därmed på grund av mer uppföljning med varje elev.

En litteraturstudie skriven av Ginsburg (2009) behandlar frågan om hur lärare kan använda kunskapen om matematiskt tänkande för att förbättra undervisningen. Studien fokuserar på lärares och elevers tankar vid formativ bedömning. Författaren menar att om läraren har en djup förståelse för hur eleverna lär sig kan undervisningen förbättras. Ginsburg (2009) anser att läraren främst behöver veta fyra olika saker om sina elever för att den formativa bedömningen ska leda till att förbättra undervisningen. Först behöver läraren veta om eleven behärskar ett visst område eller har förstått det som läraren undervisat om. Läraren måste också veta hur eleven tänker för att ta reda på vilka svagheter eller styrkor eleven har. Vidare ska läraren även ha insikt i elevens lärandepotential, alltså ha kunskap om vad eleven är redo att lära sig. Till sist menar Ginsburg (2009) att läraren måste vara intresserad av vilken motivation eleverna har. För att ta reda på detta kan läraren enligt Ginsburg (2009) använda sig av tre olika metoder, nämligen observation, prov eller uppgift och intervju med eleverna. Utefter den kunskap som läraren har om sina elever kan undervisningen byggas upp på ett sätt som främjar elevernas lärande.

Ginsburg (2009) kom fram till att läraren först bör förklara de lärandemål som ska uppnås. Utefter elevens individuella nivå kan flera olika definitioner behövas för att alla elever ska förstå vad som är målet med undervisningen. Det är även bra om läraren har förutspått vilka svårigheter eleverna kan stöta på för att kunna hjälpa dem på ett bra sätt. Läraren behöver också utvärdera sin undervisning och fundera över vad som kan förbättras för att skapa en så hög kvalitet på undervisningen som möjligt. Resultatet av den formativa bedömningen beror på många olika faktorer enligt Ginsburg (2009). Författarens slutsats är att läraren försöker förstå matematiken, sina elevers tankar och principerna för undervisningen men i slutänden handlar det om att tillämpa alla idéer på ett personligt och meningsfullt sätt för att undervisa eleverna här och nu.

Aureliana-Loredana (2014) utförde en litteraturstudie som behandlar elevers motivation för lärande med hjälp av alternativa bedömningsmetoder. Författaren har fokuserat på elever i högre utbildningar och syftar till att identifiera effektiva metoder för bedömning. Det Aureliana-Loredana (2014) ville ta reda på var i huvudsak tre förhållanden: relationen mellan

alternativa bedömningsmetoder och elevers prestationer, relationen mellan elevers motivation och den återkoppling de får samt rollen av formativ bedömning och slutbedömning för att öka elevers motivation till att lära sig. Aureliana-Loredana (2014) menar att motivation är den faktor som uppmuntrar elever att anstränga sig för att lyckas och därmed undvika att misslyckas. Författaren menar vidare att läraren är viktig för elevers motivation då läraren fungerar stöttande och har ett bra samarbete med eleven. Med alternativa bedömningsmetoder menar Aureliana-Loredana (2014) allt annat än de traditionella papper och penna prov som ofta är underlag för bedömning av elevers kunskaper. Med hjälp av portfolios anser författaren att eleven utvecklar en bättre förmåga till självbedömning och det bidrar även till ett mer aktivt deltagande i sitt eget lärande. Eleverna kan sätta upp sina egna mål och lätt se sina framgångar i en portfolio, vilket fungerar motivationshöjande. Den kontinuerliga återkopplingen från läraren bidrar också till en ökad motivation. Genom att blanda både verbal och skriven återkoppling till eleverna utvecklar det ett bättre lärande hos dem. Aureliana-Loredana (2014) menar att eleverna anstränger sig mer om de får återkoppling som både är skriftlig och verbal. Sammanfattningsvis kom författaren fram till att alternativa bedömningsmetoder är bättre än de traditionella metoderna för elevernas utveckling.

#### 5.1.1 Lärares val av bedömningsformer

Van den Heuvel-Panhuizen & Veldhuis (2014) genomförde en studie som syftade till att karaktärisera lärares olika bedömningsformer. I studien deltog 960 lärare från 557 olika skolor i Nederländerna. Data till studien samlades in genom att lärarna fick besvara en enkät som skickades ut via internet. Analysen av materialet utfördes i två steg. Det första steget gick ut på att granska enkätens struktur samt dela in lärarna i olika underklasser. Nästa steg bestod av att författarna studerade skillnader mellan de olika underklasserna av lärarna utifrån valda faktorer.

Enligt van den Heuvel-Panhuizen & Veldhuis (2014) visade resultatet att lärarna som besvarat enkäten använde sig av en blandning av observation och instrumentbaserade metoder i matematikundervisningen. Den mest använda observationsbaserade metoden var frågor, observationer och att rätta skrivet arbete. Den primära instrumentbaserade metoden var prov som behandlade samma stoff som läroboken. Lärarna använde dessa metoder som en blandning av summativa, formativa och diagnostiserande syften. Den summativa delen syftade till att välja vad som ska läras ut i matematikundervisningen. Den formativa delen avsåg att tillhandahålla återkoppling, avgöra hastigheten på undervisningen och för att



anpassa instruktioner. Det diagnosticerande syftet var att undersöka anledningar till varför saker i undervisningen går fel.

En litteraturstudie som utförts av Tveit (2013) har handlat om bedömning i Norge. Tveit (2013) tar i sin litteraturstudie upp betydelsen av formativ och summativ bedömning i undervisningen samt om nationella kriterier är ett bra verktyg att ha i undervisningen med hänsyn till de olika bedömningsformerna. Resultaten utifrån hans litteraturstudie visade bland annat att lärares bedömningar har stor betydelse för eleverna och att elever lär sig bäst när de förstår vad det är de ska lära sig och vad som förväntas av dem. Resultaten visar även att eleverna lär sig bäst när de får återkoppling som informerar dem om kvalitén på deras arbeten eller prestationer, får tips och råd om hur de kan förbättras, är involverade i sitt eget lärande genom att till exempel bedöma sina egna arbeten och utveckling. Enligt Tveit (2013) pekar den norska styrelsen av utbildning mot forskningen som föreslår att bedömning för lärande, dvs. formativ bedömning, är ett av de mest effektiva sätten att öka elevernas prestationer samt deras möjligheter för lärande. Tveit (2013) fokuserade främst på hur bedömning specifikt ser ut i just Norge och sammanfattade att det behövs både summativa och formativa bedömningsformer. Studien visade även att en blandning av dessa bedömningsformer är det som majoriteten av lärarna använder. Det är dock svårt för norska lärare att lägga lika stor vikt på båda bedömningsformerna vid till exempel nationella prov där rankingen av skolor bland annat kan överskugga syftet för formativ bedömning och fokus läggs därför mer på den summativa delen istället för den formativa delen.

## 5.2 Summering av resultat om bedömningsformer

Resultaten visar att formativ bedömning är en bedömningsmetod som lärare anser är bra och vill använda sig av i sin undervisning. Forskning visar att både lärare och elever uppskattar formativ bedömning. Det finns många fördelar med användandet av bedömning för lärande som till exempel att det främjar elevernas kunskapsutveckling, eleverna har större inflytande och insyn i sitt eget lärande samt att det ökar elevernas motivation för lärande. Forskare poängterar dock att det finns en nackdel med formativ bedömning vilket är att lärare får ökad arbetsbörda och får en större upplevelse av tidspress. För att den formativa bedömningen ska uppnå sitt syfte bör läraren bland annat ha djup förståelse för vad eleverna behärskar i ett visst område. För att lärare ska ha denna inblick beträffande varje enskild elev krävs det mycket tid, tid som lärare hellre vill lägga på lektionsplanering för att eleverna ska nå

kunskapskraven. Flera studier visar att en blandning av summativa och formativa bedömningar är det som de flesta lärare använder sig av.

### 5.3 Flervalsfrågors betydelse i matematikprov

Ventouras, Triantis, Tsiakas och Stergiopoulos (2009) har gjort en studie i en vuxenutbildning för att jämföra bedömningsmetoder för prov med flervalsfrågor och prov med frågor där eleven konstruerar sitt eget svar. Studiens undersökning bestod av ett prov där eleverna konstruerade sina egna svar och två prov med flervalsfrågor där två olika typer av bedömningsmetoder användes. Den ena bedömningsmetoden var att eleven fick en poäng för varje svar som var korrekt medan den andra innebar att det var frågor i par som behandlade samma sak. Det framkom dock inte tydligt att de handlade om samma sak ifall den examinerade inte hade de rätta kunskaperna om ämnet. 62 elever deltog vid första tillfället där de två första proven utfördes och 63 elever deltog vid det andra tillfället då det sista provet utfördes. Vid konstruerandet av varje prov togs hänsyn till att alla frågor skulle ha samma svårighetsgrad. Det författarna ville granska var de olika bedömningsmetoder som hörde till de olika sorters prov. Det första provet hade fem frågor som krävde att eleven konstruerade längre och beskrivande svar. Efter utsatt tid skickades svaren automatiskt till examinatorn och till skrivaren så att alla elever fick en kopia på hans/hennes svar. Det andra provet bestod av 40 stycken slumpvist utvalda flervalsfrågor. Efter utsatt tid visades en resultatsida för varje enskild elev där alla rätt och fel svar var synliga. Bedömningen för detta prov heter ”positiv rättning” och går ut på att endast ge poäng för korrekt besvarade frågor, inget poäng ges för fel eller uteblivet svar. Vid det andra tillfället utfördes det sista provet som bestod av parade flervalsfrågor. Det innebär att det var sammanlagt 60 frågor då 30 frågor var parvis sammansatta. Hade eleven kunskaper om ämnet skulle eleven kunna besvara båda frågorna i ett par korrekt.

Ventouras m.fl. (2009) resultat från undersökningen visade att eleverna fick högre poäng vid prov med flervalsfrågor än med frågor där de fick konstruera svaren själva. Detta metodfel var uppenbart och författarna menar att det kan vara troligt att det finns ett samband med faktorn av att ha ren tur, det vill säga att eleverna svarade rätt på frågor av ren slump. Författarnas resultat visar att flervalsfrågor och frågor med egen konstruerade svar inte är ekvivalenta, alltså inte likvärdiga. Frågorna var anpassade och formulerade så att de skulle behandla samma saker och fler elever fick ändå bättre resultat på flervalsprovet än provet med egna konstruerade svar. Det var både signifikanta skillnader mellan det första provet (egna

konstruerade svar) och det andra (flervalsfrågor) och mellan det andra provet och det tredje (parade flervalsfrågor). Författarna sammanfattade att flervalsfrågor med negativ rättning, dvs. att det dras bort poäng för varje svar som är inkorrekt besvarat, var statistiskt ekvivalenta till frågor med egen konstruerade svar. Faktorn för metodfel från ren tur som fanns för det andra provet (flervalsfrågor) försvinner vid prov med parade flervalsfrågor med negativ rättning. Författarna menar dock att det finns vissa nackdelar med negativ rättning och att elever kan hämmas av rädslan för minuspoäng och strunta i att ta sig an en fråga som den har någorlunda kunskaper om. Bedömning baserad på parade flervalsfrågor kan utgöra ett pålitligt verktyg för utvärdering och rättning av elevers prestanda. Ventouras m.fl. (2009) menar att fördelarna med bedömning baserat på flervalsfrågor, som exempelvis tiden för rättning, den rättvisa fördelningen av poäng samt att elevers prestationer i allt som tagits upp i undervisningen kan examineras och utnyttjas fullt ut så länge det är parade flervalsfrågor eller med negativ rättning.

Hastedt och Sibberns (2005) har gjort en studie där de har granskat TIMSS resultaten från år 1995 och 1999. Syftet med deras studie var att kunna värdera ifall de olika frågeformuleringarna, flervalsfrågor och frågor med egen konstruerade svar, ger olika resultat. Eftersom data endast var tillgänglig för årgruppen 15 åringar så lades granskningen på den åldern i denna studie. I TIMSS 1995 var det 45 länder som deltog med 271 995 elever som gick i årskurs 8. I den matematiska delen fanns det 124 flervalsfrågor och 35 frågor där eleverna fick konstruera sina egna svar. I TIMSS 1999 var det 38 länder som deltog med 184 863 elever från årskurs 8. I den matematiska delen år 1999 förekom det 126 flervalsfrågor och 37 frågor med egen konstruerade svar. Resultatet av granskningen visade att vissa länder gynnades av flervalsfrågor medan andra länder gynnades av frågor med egen konstruerade svar i både TIMSS 1995 och TIMSS 1999. Ett exempel på detta är Irland och Slovenien där Irland presterade betydligt bättre på delen med egen konstruerade svar än Slovenien medan Slovenien presterade betydligt mycket bättre än Irland på flervalsfrågorna. Författarna refererar också till forskning som lägger vikt på att flickor oftast presterar bättre på frågor med egen konstruerade svar och pojkar presterar bättre på prov med flervalsfrågor. Detta bekräftades i Hastedt och Sibberns (2005) studie då deras granskning visade att i de allra flesta fall var det flickor som fick högre resultat på frågor med egen konstruerade svar och pojkar presterade bättre på flervalsfrågor. Det förekom enstaka fall, som till exempel i Colombia i TIMSS 1995, där pojkarna hade presterat bättre på både frågor med egen konstruerade svar och flervalsfrågor. Hastedt och Sibberns (2005) lägger dock vikt på att det

finns ett behov för mer forskning i det här ämnet. Sammanfattningsvis avslutar Hastedt och Sibberns (2005) med att ge slutsatsen av att formuleringen av frågor i ett prov kan påverka resultaten.

Roediger III och Marsh (2005) har gjort en studie där de undersökte konsekvenserna med prov bestående av flervalsfrågor. Studien utfördes genom att studenterna fick läsa olika stycken ur en bok inför ett prov med flervalsfrågor. De fick 18 stycken att läsa utifrån 36 stycken som fanns, så de fick inte läsa på om allt som skulle examineras. Det var 24 studenter från Washington University som deltog. Provet bestod av 144 flervalsfrågor och efteråt skulle studenterna besvara ett så kallat "cued-recall test" vilket är ett slags minnesprov för att se hur mycket elever kan komma ihåg från till exempel en kurs eller ett prov. Minnesprovet bestod av 216 frågor. I provet med flervalsfrågorna var studenterna uppmanade att svara på varje fråga. Det förekom ingen återkoppling efter proven så studenterna fick inte veta ifall de hade besvarat frågorna rätt eller fel. Studenterna blev också informerade vid provtillfället att det kunde finnas flera svar som var korrekta på vissa frågor. På minnesprovet blev studenterna uppmanade att endast svara på de frågor som de var någorlunda säkra på och blev därmed varnade för att inte gissa. Resultaten av proven visade att flera av de frågorna utifrån de styckena som studenterna fick läsa på innan provet var korrekt besvarade. Studenterna svarade också korrekt på fler frågor när det fanns färre alternativ eftersom det fanns större chans att gissa rätt. När det kom till frågor baserade på de förlästa styckena svarade eleverna rätt på fler frågor med två alternativ än med fyra, men de presterade inte sämre vid frågor med sex alternativ än med fyra. Det syntes sämre prestationer vid frågor som var baserade på styckena som studenterna inte fick läsa innan när svarsalternativen ökades från fyra till sex. Författarnas slutsats var att studenterna presterade sämre vid prov med flervalsfrågor när det fanns fler alternativ och där det fanns material som var obekant.

Burton (2004) har gjort en studie där han diskuterar och testat om det går att mäta reliabiliteten i prov med sant/falskt- och flervalsfrågor. Författaren går in på olika bedömningsmetoder för sådana här prov som till exempel "negative marking" (negativ rättning) där bedömare ger poäng för de korrekt besvarade frågorna och minuspoäng för de svaren som besvarats inkorrekt. Burton (2004) nämner hur negativ rättning, när alla elever har förstått hela innebörden av metoden, förbättrar reliabiliteten i ett prov med flervalsfrågor eftersom det avskräcker eleverna från att gissa. Författaren lägger dessutom tyngd på att genom att enbart gissa på ett prov med flervalsfrågor där alla frågor har fyra svarsalternativ

kan det resultera i 25 % rätt och i sant/falskt frågor kan eleven få upp till 50 % rätt genom att enbart gissa. Bedömaren bör dock skilja på svar som är allmänt gissade på och svar där eleven har haft delvisa kunskaper när den har gissat (Burton, 2004). I studien prövade Burton (2004) en uträkning som sägs kunna mäta reliabiliteten i ett prov med flervalsfrågor och sant/falskt frågor, han kommer dock fram till att det går inte med en enkel uträkning att bestämma reliabiliteten i ett prov då bedömaren måste ta hänsyn till olika faktorer som kan påverka ett prov. Antalet svar som är besvarade genom gissningar beror helt på elevernas personlighet, villkor vid ett prov och förväntningar av betyg enligt Burton (2004). Författarens slutsats var att vid planering av ett prov kan ett provs reliabilitet bedömas i termer av bland annat modellbaserade mätningar, med hänsyn till provlängden och antal svarsalternativ per fråga. Hur reliabiliteten bäst kan fastställas när testdata finns tillgänglig beror både på varför kvantifieringen är nödvändig och om vad provet är tänkt att mäta. Detta beror på att det, som nämnt tidigare, inte finns något generellt sätt att sammanfatta reliabiliteten i ett prov genom en enda uträkning eftersom det behövs ytterligare information (Burton, 2004). Slutligen lägger författaren vikt vid att den information eleverna bär med sig till examinationerna förtjänar mer uppmärksamhet i förhållande till reliabiliteten i ett prov med sant/falskt frågor eller flervalsfrågor. Den informationen kan till exempel vara att få negativa rättningar vid felaktigt besvarade frågor. Burton (2004) menar att när eleverna väl är medvetna om konsekvenserna av gissning i samband med negativ rättning ökar reliabiliteten i matematikprovet.

Lau, Lau, Hong och Usop (2011) har också utfört en studie som handlar om flervalsfrågor och i vilken grad eleverna gissar samt hur läraren kan ta reda på om eleverna har någon kunskap inom ämnet. Studien utfördes i Malaysia där bedömning oftast sker genom att använda flervalsfrågor vid provtillfällen. Studien hade 449 deltagande elever i åldrarna 13-14 år från 19 olika skolor och genomfördes i matematikämnet. Författarna jämförde två olika datorprogram som båda har flervalsfrågor vid prov men som kan bedömas på olika sätt. Det första programmet består av frågor där eleven väljer ett svarsalternativ som han/hon tror är rätt och går sedan vidare till nästa fråga. Det andra, mer välutvecklade, programmet består av frågor som ska besvaras genom att både välja vad som är fel och vad som är rätt. I det programmet finns svarsalternativen ”rätt”, ”fel” och ”jag vet inte”. Anledningen till att det finns fler svarsalternativ är för att upptäcka om eleven har missuppfattat en fråga eller har delvis kunskap inom området. Poängsättningen ser lite annorlunda ut till skillnad från det första programmet, där varje korrekt besvarad fråga ger ett poäng och om eleven svarat fel

eller inte gett något svar alls blir inga poäng utdelade för frågan. I det andra programmet får eleverna ett poäng för varje korrekt ”fel” de väljer, de får ett poäng vid rätt svar och om eleven väljer ”fel” vid rätt svar får de tre minuspoäng. Alternativet ”jag vet inte” ger varken plus- eller minuspoäng. Anledningen till att författarna valde att ha med minuspoäng var för att förhindra gissningar.

De tre frågor som studien syftade till att ta reda på var i vilken utsträckning elever gissar, om det nya programmet kan upptäcka delvis kunskap och missuppfattningar samt om det nya programmet har en högre reliabilitet. Resultatet visar att gissning blir minimalt med det nya programmet eftersom det kan leda till minuspoäng. I det äldre programmet var det svårt att veta om eleverna gissat eller inte men i det nya programmet blir det mer tydligt och eleverna kan välja alternativet ”jag vet inte” om de är osäkra. Det var även lättare att upptäcka om eleverna missuppfattat en fråga eller hade lite kunskap om en fråga. Tack vare de olika svarsalternativen kunde de delas in i fem olika grupper istället för två som det tidigare programmet hade, nämligen ”rätt” och ”fel”. Det nya programmet hade istället grupperna ”total missuppfattning”, ”delvis missuppfattning”, ”ingen kunskap”, ”delvis kunskap” och ”total kunskap”. Detta gjorde det tydligt för lärarna att se vad de behövde undervisa mer om, om flera elever hade problem vid samma typ av fråga. Reliabiliteten ökade vid användandet av det nya programmet, vilket innebär att tillförlitligheten är högre. I det nya programmet visar alla positiva poäng att eleverna har någon kunskap och vid minuspoäng visar det att eleverna har missuppfattat någonting. Lau, m.fl. (2011) tar dock upp att det finns andra faktorer som kan påverka ett resultat på ett prov. Några av dessa faktorer menar författarna kan vara provtagarens beteende, administration eller provsituation. Sammanfattningsvis kan dock sägas att det nya programmet har en högre reliabilitet än det gamla och att det förhindrar gissningar.

#### 5.4 Summering av resultat om skriftliga frågor

Resultaten i vår litteraturstudie kring reliabiliteten i prov med flervalsfrågor har visat att det finns olika faktorer som påverkar reliabiliteten i ett prov med flervalsfrågor som exempelvis att eleverna kan gissa sig fram till rätt svar. Det går dock att öka reliabiliteten genom att till exempel använda sig av parade flervalsfrågor eller negativ rättning, där eleverna får minuspoäng för varje felaktigt besvarad fråga. Forskarna poängterar dock att det finns en nackdel med negativ rättning och det är att eleverna kan hämmas av rädslan för minuspoäng

och därmed strunta i att ta sig an en fråga som eleven ändå har någorlunda kunskaper om. Forskningen visar att flickor presterar i snitt bättre än pojkar vid prov med frågor där eleverna får konstruera svaren själva medan pojkar presterar bättre än flickor på prov med flervalsfrågor. Forskarna påpekar också att tittar man internationellt visar det sig att vissa länder gynnas av prov med flervalsfrågor medan andra länder missgynnas av det och presterar betydligt bättre vid prov med frågor där eleverna konstruerar svaren själva.

## 6.0 Diskussion

Denna studie har gått ut på att undersöka hur olika bedömningsformer främjar elevernas matematiklärande samt hur reliabiliteten i ett matematikprov påverkas av flervalsfrågor. Syftet är att uppmärksamma betydande faktorer inom bedömning för nuvarande och blivande lärare. I följande avsnitt utförs en analys och diskussion om litteraturstudiens resultat med koppling till studiens bakgrund. Därefter följer ett avsnitt om faktorer som påverkar resultaten och sedan presenteras en slutsats.

### 6.1 Formativ bedömning påverkar elevers motivation

Aureliana-Loredana (2014) lägger vikt på elevers motivation i undervisningen och menar att motivation är den faktor som uppmuntrar elever att anstränga sig för att lyckas. Om eleverna är motiverade för att lära sig nya saker kommer de ha lättare att lära sig. Aureliana-Loredana (2014) nämner i sin studie att läraren är viktig för elevers motivation då läraren fungerar stöttande och har ett bra samarbete med eleverna. Jonsson, Lundahl och Holmgren (2015) nämner utifrån deras studie att elevers motivation faktiskt höjdes på grund av den formativa bedömningen de fick från sina lärare. Tidigare i arbetet nämner vi självbedömning och kamratbedömning som två typer av formativ bedömning där eleverna är mer involverade i bedömningsprocessen. Genom att använda självbedömning och kamratbedömning i undervisningen sparar läraren tid eftersom den formativa bedömningen sker mellan eleverna (Woolfolk, 2014). Aureliana-Loredana (2014) menar även att självbedömning kan verka motivationshöjande för eleverna då de enkelt kan se sina prestationer och framsteg. Läroplanen (Skolverket, 2011) tar upp att målet i all undervisning är att utveckla intresse, nyfikenhet och förmågor hos alla elever så att de därmed lär sig grundläggande kunskaper samt att bedömningskulturen i skolan spelar stor roll för elevers inställning till lärande.

### 6.2 Bedömningsformer och läraren

Av resultatet framgår att formativ bedömning är en bedömningsmetod som många lärare är nyfikna på och vill använda i sin undervisning, men de har inte alltid förutsättningarna eller stödet för att utföra detta. Woolfolk (2014) menar att formativ bedömning är en typ av bedömning som syftar till att stärka elevers lärande. Resultaten i vår litteraturstudie visar att formativ bedömning verkligen gör detta. En nackdel är dock att den formativa bedömningen förutsätter att läraren är mer fördjupad i varje enskild elevs kunskapsutveckling vilket kräver mycket av lärarens tid (Antoniou & James, 2014; Ginsburg, 2009; Jonsson m.fl., 2015;



Aureliana-Loredana, 2014; van den Heuvel-Panhuizen & Veldhuis, 2014; Tveit, 2013). Det innebär att det blir en ökad arbetsbörda för lärarna då det blir avsevärt mer att göra under arbetstiden. Ginsburg (2009) nämner att det finns en rad olika saker som läraren behöver ha kunskaper om och detta kräver att läraren lägger mer tid på att tillägna sig dessa kunskaper. Den bästa bedömningsmetoden är formativ bedömning då den bidrar till elevernas matematikutveckling medan den summativa bedömningen ger ett svar om hur eleven presterade vid just det bedömningstillfället, oftast kort som till exempel ”Bra jobbat!”. Eleven får på det sättet inte reda på vad den kan förbättra eller hur de kan förbättras. Den summativa bedömningen handlar mer om att summera elevens kunskaper. Nackdelen med formativ bedömning är en förutsättning som inte riktigt går att ändra idag (Ginsburg, 2009). Utifrån våra erfarenheter har vi märkt att lärare redan känner sig pressade och ofta får jobba övertid. Formativ bedömning är ändå en bedömningsform som det bör läggas mycket fokus på i skolan och lärare bör lägga så mycket tid som möjligt för att utföra formativa bedömningar då det främjar elevernas kunskapsutveckling.

Tveits (2013) litteraturstudie visade att det var utmanande för lärare att använda en blandning av summativa och formativa bedömningar då det lätt blev att den formativa bedömningen överskuggades av den summativa bedömningen på grund av t.ex. nationella rankingar. Utifrån vårt resultat anser vi att lärare bör lägga så mycket tid de kan på formativa bedömningar och planerar lärare sin arbetstid på rätt sätt kan de eventuellt få mer tid till det. Observationer sker kontinuerligt under skolans gång och denna bedömningsform är någonting lärare använder sig av hela tiden, utan att ens tänka på att de gör det ibland. Utifrån sin observation kan läraren få information för formativa bedömningar och vi tror utifrån våra erfarenheter att vissa lärare observerar och tar in informationen men går inte vidare med att förmedla informationen till eleverna. Skolverket (2010) nämner att den formativa bedömningen karakteriseras av att läraren klargör undervisningens mål för eleverna, vet var eleverna befinner sig i förhållande till målen och att eleverna får återkoppling som syftar till hur målen ska uppnås. En formativ bedömning gynnar eleverna betydligt mycket mer än en summativ bedömning men eftersom tidsbrist finns föreslår vi utifrån studiens resultat att en blandning av summativa och formativa bedömningar är det bästa metoden för lärare och elever. Van den Heuvel-Panhuizen och Veldhuis (2014) studie visar även att en blandning av summativa och formativa bedömningar är den metod de flesta lärare använder. Enbart summativa bedömningar skulle inte gynna eleverna alls medan enbart formativa bedömningar kräver alldeles för mycket tid av lärarna och därför resonerar vi som vi gör. Formativa bedömningar behövs i

undervisningen men ibland är det kanske tillräckligt med en summativ bedömning. Antoniou och James (2014) studie ledde till att en av de deltagande lärarna ändrade sin undervisning efter att ha tolkat elevernas lärande och var den enda som tog till sig konceptet av formativ bedömning. Detta visar att det inte är omöjligt för lärare att anpassa sin undervisning utifrån elevernas lärande.

### 6.3 Återkoppling är bra för eleverna

Jonsson m.fl. (2015) samt Antoniou och James (2014) nämner att återkoppling är en viktig del i arbetet med att bedöma eleverna och detta anser även Skolverket (2011d). I artiklarna talar författarna om att ge eleverna konstruktiv feedback för att reglera elevernas lärande. Det handlar om att veta vart eleven befinner sig, vart eleven ska och hur eleven ska göra för att nå målen (Jonsson m.fl. 2015). Jonsson m.fl. (2015) menar vidare att dokumentation, uppföljning och kommunikation med eleverna om deras framsteg är viktigt och bidrar till högre kvalitet på elevernas arbeten. Aureliana-Loredana (2014) menar också att regelbunden återkoppling leder till ett bättre lärande hos eleverna. Författaren anser att en blandning av verbal och skriftlig återkoppling gör att eleverna anstränger sig mer. Tveit (2013) håller med om dessa påståenden och menar att återkoppling ska informera eleverna om kvaliteten på deras arbeten samt bidra med tips på hur eleverna kan förbättras. Utifrån våra egna erfarenheter från verksamhetsförlagd utbildning har vi märkt att lärare ofta ser återkoppling som något han/hon själv måste ge eleverna. Återkoppling är en del av den formativa bedömningen och som nämnts tidigare kräver det mycket av lärarens tid. Vi vill gärna framhäva att det finns andra metoder för återkoppling som inte kräver lika mycket tid.

Tveit (2013) påpekar att eleverna utvecklas av självbedömning och kamratbedömning. Eleverna blir mer involverade i sitt eget lärande med denna typ av bedömning vilket leder till personlig utveckling (Tveit, 2013). Även Jonsson m.fl. (2015) framhåller vikten av att aktivera eleverna som ägare av sitt eget lärande och som resurser för varandras lärande. Författarna menar vidare att användning av bedömningsmatriser kan leda till att eleverna blir mer medvetna om vilka mål de ska uppnå och hur de ska ta sig dit. Aureliana-Loredana (2014) framhäver istället portfolios som ett verktyg för självbedömning. Både portfolios och bedömningsmatriser kan användas för att göra eleverna mer uppmärksamma på vad de arbetar mot för mål. Aureliana-Loredanas (2014) studie fokuserade på äldre åldrar vilket innebär att portfolios kanske inte är lika användbara i lågstadiet som i högre utbildningar. I lägre åldrar skulle det kanske krävas att läraren kontrollerar att alla elever använder sin portfolio då

eleverna behöver mer vägledning från läraren eftersom det som ska finnas i portfolion annars kan hamna på fel ställe. Det kanske dock inte är fallet om läraren kan anpassa användandet av portfolios till lägre åldrar.

Återkoppling ska användas med syfte att öka elevernas lärande (Aureliana-Loredana, 2014). Det syftar också till att läraren anpassar sina instruktioner efter vad eleverna behärskar och kan även hjälpa läraren att avgöra om hastigheten på undervisningen fungerar bra eller inte (van den Heuvel-Panhuizen & Veldhuis, 2014). Vi anser därför att återkoppling är bra om den sker på rätt sätt och att det inte alltid innebär att läraren förmedlar den till eleverna. Att bara skriva ”Bra jobbat!” eller ge en viss poäng på ett prov hjälper inte eleverna vidare i sin utveckling.

#### 6.4 Flervalsfrågor i prov i lågstadiet

Resultatet i denna studie visade att prov med flervalsfrågor inte är ekvivalenta med andra sorters prov som till exempel prov med öppna/slutna frågor. Det leder till att resultaten i PISA och TIMSS undersökningarna kan anses vara opålitliga eftersom de använder flervalsfrågor (Skolverket, 2012; Skolverket, 2014c). Vid bedömningstillfällen med flervalsfrågor kan inte läraren avgöra om eleverna kan svaren eller om de har gissat. Burton (2004) kom i sin studie fram till att det inte går att avgöra reliabiliteten i ett prov med flervalsfrågor eftersom varje enskilt fall inte är sig likt. Det går inte att förutspå hur många frågor eleverna gissar på och hur många de kan svaret på och varje examinationstillfälle skiljer sig från nästa. Roediger III och Marsh (2005) märkte i sin studie att eleverna besvarade fler frågor korrekt när det fanns färre svarsalternativ.

Vid en fråga med två alternativ, till exempel sant/falsk frågor, finns det 50 % chans att eleven svarar korrekt utan några kunskaper om ämnet. Vid en flervalsfråga med fyra svarsalternativ finns det 25 % chans att eleverna svarar rätt på en fråga genom att enbart gissa. Roediger III och Marsh (2005) har bekräftat denna faktor med sin studie eftersom det visade sig att ju färre svarsalternativ, desto större chans att eleven kan gissa rätt svar. Därför bör flervalsfrågorna ha minst fyra svarsalternativ för att öka reliabiliteten. Roediger III och Marsh (2005) nämnde att det inte var någon skillnad om det fanns fyra eller sex svarsalternativ. Detta är för att om eleven inte har det rätta svaret så får den analysera alla svarsalternativ och sedan utgå från dess delvisa kunskap och försöka besvara det som den tror är rätt (Lau m.fl., 2011). På det sättet kan läraren ändå bedöma att eleven har delvisa kunskaper. Ventouras m.fl. (2009) kom

fram till att parade flervalsfrågor eller prov med flervalsfrågor som rättades med negativ rättning hade högre reliabilitet än prov med vanliga flervalsfrågor med positiv rättning.

Woolfolk (2014) menar att det blir högre validitet när flera prov mäter samma kunskaper. Parade flervalsfrågor höjer därför validiteten och reliabiliteten i prov. Ett annat sätt att höja reliabiliteten är att använda sig av test-retest metoden (Woolfolk, 2014). Hastedt och Sibberns (2005) studie visade att flervalsfrågor gynnar elever i vissa länder medan elever i andra länder gynnas av prov med öppna frågor. Det kom också fram att flickor presterar bättre vid öppna frågor medan pojkar presterar bättre vid flervalsfrågor. Varför det skiljer sig så markant framgår inte men eftersom gissning inte riktigt är möjligt vid öppna frågor bör reliabiliteten vara högre i ett sådant prov än flervalsprovet.

Parade flervalsfrågor och negativ rättning vid flervalsfrågor nämns av flera forskare i resultatet som en användbar metod (Ventouras m.fl., 2009; Hastedt & Sibberns, 2005; Roediger III & Marsh, 2005; Burton, 2004; Lau m.fl., 2011). Det finns fördelar vid prov med flervalsfrågor som till exempel att läraren besparar tid och att det blir en rättvisare bedömning jämfört med frågor där eleverna konstruerar svaren själva (Ventouras m.fl., 2009). Eftersom det finns vissa fördelar med prov med flervalsfrågor anser vi att om läraren höjer reliabiliteten genom att till exempel använda sig av parade flervalsfrågor och/eller negativ rättning samt fler svarsalternativ kan denna typ av fråga användas. Nackdelarna som förekommer är som nämnt att elever har möjlighet att gissa sig fram till det rätta svaret och att läraren bedömer endast elevens kompetens i att ge ett rätt svar och inte bedömer hur eleven kom fram till svaret. Fördelen med öppna frågor är att eleven visar konkret hur den har tänkt, oavsett om svaret är rätt eller fel. Genom prov med flervalsfrågor kan läraren endast utföra en summativ bedömning medan andra sorters prov, som till exempel prov med öppna frågor kan bidra till en formativ bedömning eftersom läraren kan se hur eleven har tänkt och resonerat. Ventouras m.fl. (2009) skriver att det finns nackdelar vid prov med öppna frågor som till exempel att det tar längre tid att rätta samt att alla svar är individuella vilket gör det svårare för läraren att ge en objektiv och rättvis bedömning.

Forskningen som presenteras i resultatet har dock inte utförts i lågstadiet, vilket vi även noterat i metoddiskussionen. Detta resulterar i att vi inte kan dra några slutsatser att samma resultat skulle visa sig om forskningen hade utförts i lågstadiet. Anledningen till varför vi ändå valde att använda den forskningen är delvis på grund av den begränsade tidsramen, saknaden av forskning i lågstadiet och också på grund av att elevers möjlighet att kunna gissa

på ett prov med flervalsfrågor finns oavsett vilken ålder de befinner sig i. Det som skiljer är dock att prov med öppna frågor i lågstadiets matematikundervisning eventuellt inte innehåller krav på särskilt utvecklade svar. Detta gör att det enda som skiljer sig mellan de olika frågeformuleringarna är möjligheten att gissa på frågor. Frågor där det endast finns en tom rad för ett svar kräver att eleven svarar själv medan flervalsfrågor ändå exponerar eleven för olika svar där eleven vet att ett av dessa alternativ är det korrekta.

## 7.0 Slutsats

Sammanfattningsvis kan vi konstatera att formativ bedömning främjar elevers matematikutveckling. Denna typ av bedömning kräver dock mycket tid av läraren och ökar därmed arbetsbördan. Därför är en blandning av formativa och summativa bedömningar det som främjar både eleverna och läraren. Den formativa delen är den som främjar eleverna mest medan den summativa delen underlättar lärarens arbete. När läraren ska göra en formativ bedömning utifrån ett matematikprov är flervalsfrågor inte att föredra. Flervalsfrågor har låg reliabilitet eftersom möjligheten att gissa finns och läraren kan inte bedöma elevernas sanna kunskaper. Det finns däremot vissa åtgärder som kan vidtas för att höja reliabiliteten, som exempelvis att använda negativ rättning.

## 8.0 Förslag på vidare forskning

Vår studie har gett oss resultat som besvarar våra två frågeställningar. Angående frågeställningen ”Hur påverkas reliabiliteten i ett matematikprov av flervalsfrågor?” fann vi inte mycket ny forskning som har utförts i lågstadiet som berör matematikämnet. Vi fann dock forskning om frågeställningen från till exempel 70-talet, men vi anser att det här ämnet behöver en bredare och nyare forskning med inriktning på just lågstadiet. Då våra resultat visar att flervalsfrågor påverkar reliabiliteten i ett matematikprov skulle vi gärna vilja veta hur frågeformuleringen påverkar prov i lågstadiet, då våra resultat enbart går in på äldre åldrar. Även om det inte var så svårt att hitta forskning om den första frågeställningen föreslår vi ändå att mer forskning kan utföras inom ämnet. Vi anser att ju mer forskning som framhäver skillnader mellan olika bedömningsmetoder desto tydligare blir resultatet.

## 9.0 Referenslista

\*Antoniou, P. & James, M. (2014). Exploring formative assessment in primary school classrooms: Developing a framework of actions and strategies. *Educ Asse Eval Acc.* 26, 153-176. DOI: 10.1007/s11092-013-9188-4

\*Aureliana-Loredana, P. (2014). Stimulating learning motivation of early education and primary school students through alternative methods of assessment. *Journal Plus Education. Special issue.* 303-312. Tillgänglig på: <http://www.uav.ro/jour/index.php/jpe/article/view/289>

Björklund Boistrup, L. (2010). *Assessment discourses in mathematics classrooms [Elektronisk resurs] : a multimodal social semiotic study.* Diss. Stockholm : Stockholms universitet, 2010. Stockholm. Tillgänglig på: <http://www.diva-portal.org/smash/get/diva2:355024/FULLTEXT02.pdf>

\*Burton F, Richard. (2004). Multiple-choice and true/false tests: reliability measures and some implications of negative marking. *Assessment & Evaluation in Higher Education.* 29(5). DOI: 10.1080/02602930410001689153

Eriksson Barajas, K., Forsberg, C. & Wengström, Y. (2013). *Systematiska litteraturstudier i utbildningsvetenskap: vägledning vid examensarbeten och vetenskapliga artiklar.* (1. utg.) Stockholm: Natur & Kultur.

\*Ginsburg, H. (2009). The Challenge of Formative Assessment in Mathematics Education: Children's Minds, Teachers' Minds. *Human Development.* 52, 109-128. DOI: 10.1159/000202729

Hartman, S.G. (2003). *Skrivhandledning för examensarbeten och rapporter.* (1. utg.) Stockholm: Natur och kultur.

\*Hastedt, D. & Sibberns, H. (2005). Differences between multiple choice items and constructed response items in the IEA TIMSS surveys. *Studies in Educational Evaluation.* 31, 145-161. DOI: 10.1016/j.stueduc.2005.05.007

\*Jonsson, A., Lundahl, C. & Holmgren, A. (2015). Evaluating a large-scale implementation of Assessment for Learning in Sweden. *Assessment in Education: Principles, Policy & Practice.* 22(1), 104-121. DOI: 10.1080/0969594X.2014.970612

Johansson, B. (2001). *Hög tid för matematik*. Göteborg: Nationellt centrum för matematikutbildning, Göteborgs univ. Tillgänglig på:

[http://ncm.gu.se/media/ncm/kup/Hog\\_tid\\_for\\_matematik.pdf](http://ncm.gu.se/media/ncm/kup/Hog_tid_for_matematik.pdf)

\*Lau, P. N. K., Lau, S. H., Hong, K. S. & Usop, H. (2011). Guessing, Partial Knowledge and Misconceptions in Multiple-Choice Tests. *Educational Technology & Society*. 14(4). 99-110. Tillgänglig på: [http://www.ifets.info/journals/14\\_4/10.pdf](http://www.ifets.info/journals/14_4/10.pdf)

Löwing, M. & Kilborn, W. (2002). *Baskunskaper i matematik: för skola, hem och samhälle*. Lund: Studentlitteratur.

\*Roediger III, H. L. & Marsh, E. J. (2005). The positive and Negative Consequences of Multiple-Choice Testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 31(5). 1155-1159. DOI: 10.1037/0278-7393.31.5.1155

Roos, J-E. & Kiselman, C.O. Nationalencyklopedin. (2015). *Matematik*, Hämtad 29 januari, 2015, från Ne.se, [www.ne.se/uppslagsverk/encyklopedi/lang/matematik](http://www.ne.se/uppslagsverk/encyklopedi/lang/matematik)

Skolverket, (2010). *Formativ bedömning – bedömning för lärande*. Hämtad 13 februari, 2015 från Skolverket,

<http://www.skolverket.se/skolutveckling/forskning/bedomning/undervisning/formativ-bedomning-1.100681>

Skolverket (2011a). *Läroplan för grundskolan, förskoleklassen och fritidshemmet 2011*. Stockholm: Skolverket.

Skolverket, (2011b). *Vad är bedömning?* Hämtad 13 februari, 2015, från Skolverket, <http://www.skolverket.se/skolutveckling/forskning/bedomning/tema-bedomning/vad-ar-bedomning-1.157698>

Skolverket, (2011c). *Vilka typer av bedömningsformer finns?* Hämtad 13 februari, 2015, från Skolverket,

<http://www.skolverket.se/skolutveckling/forskning/bedomning/tema-bedomning/vilka-typer-av-bedomningsformer-finns-1.157725>

Skolverket, (2011d). *Hur kan bedömning användas som ett verktyg för lärande?* Hämtad 13 februari, 2015 från Skolverket,

<http://www.skolverket.se/skolutveckling/forskning/bedomning/tema-bedomning/hur-kan-bedomning-anvandas-som-ett-verktyg-for-inlarning-1.157708>



Skolverket, (2012). *Fjärdeklassare har blivit bättre i naturvetenskap*. Hämtad 13 februari, 2015, från Skolverket, <http://www.skolverket.se/statistik-och-utvardering/internationella-studier/timss/fjardeklassare-har-blivit-bättre-i-naturvetenskap-1.84871>

Skolverket, (2014a). *Kraftig försämring i PISA*. Hämtad 16 februari, 2015, från Skolverket, <http://www.skolverket.se/statistik-och-utvardering/internationella-studier/pisa/kraftig-forsamring-i-pisa-1.167616>

Skolverket, (2014b). *Formativ bedömning*. Hämtad 16 februari, 2015, från Skolverket, <http://www.skolverket.se/bedomning/bedomning/formativ-bedomning-1.223359>

Skolverket, (2014c). *PISA i korthet*. Hämtad 16 februari, 2015, från Skolverket, [www.skolverket.se/statistik-och-utvardering/internationella-studier/pisa](http://www.skolverket.se/statistik-och-utvardering/internationella-studier/pisa)

Skolverket, (2014d). *TIMSS 2011. Uppgifter i matematik, årskurs 4*. Hämtad 16 februari, 2015, från Skolverket, [http://www.skolverket.se/om-skolverket/publikationer/visa-enskild-publikation?\\_xurl=http%3A%2F%2Fwww5.skolverket.se%2Fwtpub%2Fws%2Fskolbok%2Fwtpubext%2Ftrycksak%2FRecord%3Fk%3D3266](http://www.skolverket.se/om-skolverket/publikationer/visa-enskild-publikation?_xurl=http%3A%2F%2Fwww5.skolverket.se%2Fwtpub%2Fws%2Fskolbok%2Fwtpubext%2Ftrycksak%2FRecord%3Fk%3D3266)

Skolverket, (2015). *Kryss i en ruta visar inte vad eleven kan utveckla*. Hämtad 13 februari, 2015, från Skolverket, <http://www.skolverket.se/skolutveckling/forskning/bedomning/undervisning/kryss-i-en-ruta-visar-inte-vad-eleven-kan-utveckla-1.229564>

\*Tveit, S. (2013). Educational assessment in Norway. *Assessment in Education: Principles, Policy & Practice*. 21(2), 221-237. DOI: 10.1080/0969594X.2013.830079

\*Veldhuis, M. & van den Heuvel-Panhuizen, M. (2014). Primary School Teachers' Assessment Profiles in Mathematics Education. *PLoS ONE*. 9(1) DOI: 10.1371/journal.pone.0086817

\*Ventouras, E., Triantis, D., Tsiakas, P. & Stergiopoulos, C. (2009). Comparison of examination methods based on multiple-choice questions and constructed-response questions using personal computers. *Computers & Education*. 54. 455-461. DOI: 10.1016/j.compedu.2009.08.028

Woolfolk, A. (2014). *Educational psychology*. (12. utg.) Harlow: Pearson Education.

Woolfolk, A., Hughes, M., Walkup, V. & McMillan, J.H. (red.) (2011). *Educational psychology and classroom assessment*. (11. utg.) Harlow: Pearson Education.

## Bilaga 1

### *Reflektionsdokument*

Under vårt arbete med konsumtionsuppsatsen har vi tillsammans kommit fram till ett ämne vi båda är intresserade av och vill bidra med mer kunskap om. Vi har därefter utformat ett syfte med tillhörande frågeställningar samt skrivit en bakgrund till studien. Sökning, bearbetning och behandling av artiklar har vi gjort tillsammans. Diskussioner kring artiklars relevans för studien har förts mellan oss genomgående i arbetet. Vi har tillsammans skrivit resultat- och diskussionsdelen. Skrivprocessen har bestått av att den ena skriver på datorn medan den andra aktivt hjälper till att formulera texten. Vi har båda varit delaktiga och insatta i alla delar i studien och vi har alltid sett till att båda är överrens om vad som skrivs.