



SAHLGRENKA AKADEMIN

Institutionen för neurovetenskap och fysiologi  
Sektionen för hälsa och rehabilitering  
Enheten för logopedi

**350**

## **Stavningsvårigheter och taligenkänning – verktygets påverkan på barns texter**

Celia Wik  
Petter Åström

Examensarbete i logopedi,  
30 högskolepoäng  
Vårterminen 2021

Handledare  
Ingrid Henriksson  
Sanna Kraft  
Åsa Wengelin

## **Stavningsvårigheter och taligenkänning – verktygets påverkan på barns texter**

Celia Wik  
Petter Åström

*Sammanfattning.* Skrivande är komplext och utmanande för många barn. En orsak till skrivsvårigheter är att stavningssvårigheter begränsar möjligheten att använda och utveckla förmågor som behövs för att uppnå god textkvalitet. Ett verktyg som förmodats underlätta skrivandet vid stavningssvårigheter är taligenkänning men få studier har undersökt detta. I föreliggande studie undersöktes hur textkvalitet och språkliga egenskaper skiljde sig mellan tangentbordsskrivna texter och texter skrivna med taligenkänning av 58 barn med, respektive utan, stavningssvårigheter i åldrarna 10–13 år. Inga signifikanta skillnader i textkvalitet eller språkliga egenskaper fanns mellan skrivsätten för någon av grupperna. Att taligenkänning inte underlättade skrivandet för gruppen med stavningssvårigheter kan bero på ovana vid skrivverktyget eller att verktyget medför utmaningar som barnen saknade strategier för att hantera. Det fanns en stor individuell variation i hur texterna påverkades av skrivsätt, vilket antyder att andra faktorer än stavningsförmåga påverkar huruvida taligenkänning är till hjälp. Ytterligare studier behövs för att undersöka detta.

Nyckelord: skrivsvårigheter, stavningssvårigheter, assisterande verktyg, taligenkänning, textkvalitet.

## **Spelling Difficulties and Speech Recognition – Effects on Texts Produced by Children**

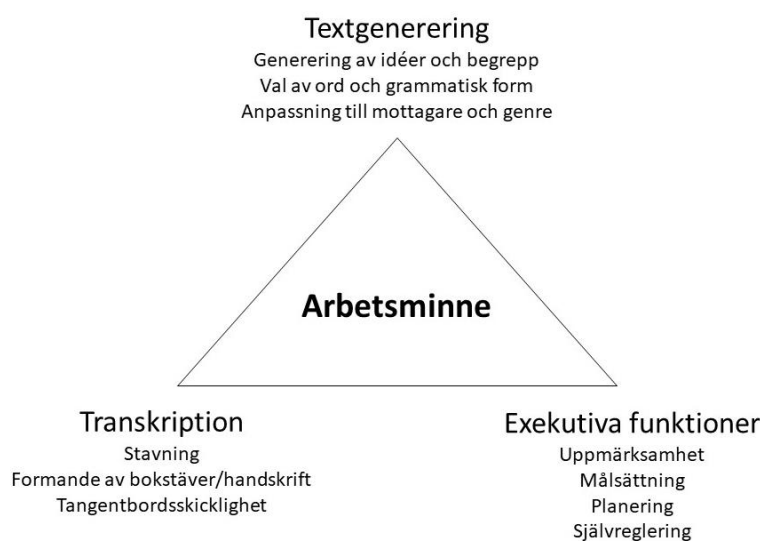
*Abstract.* Writing is a complex task that many children struggle with. Spelling difficulties might constrain use and development of high-level skills needed for writing high-quality texts. Speech recognition technology could presumably reduce these constraints for children with spelling difficulties. However, previous research is scarce and inconclusive. In the present study, text quality and linguistic features of texts produced by means of typing and speech recognition were compared. Fifty-eight children aged between 10-13 with and without spelling difficulties participated in the study. No significant effects were found for input modality for either of the groups. These results could possibly be explained by lack of experience with speech recognition and lack of strategies necessary for handling challenges imposed by the tool. More research is needed to determine whether successful use of speech recognition technology depend on individual factors, instructional practices, or both.

Key words: writing difficulties, spelling difficulties, assistive technology, speech recognition, text quality.

Skrivande är en viktig färdighet i ett allt mer digitaliserat samhälle. Kommunikation via skrift sker i utbildning, arbetsliv, i offentliga miljöer och på internet och att kunna uttrycka sig i skrift anses vara viktigt för deltagande i en rad olika aktiviteter och praktiker. Skrivutvecklingen inleds i regel redan i förskoleåldern (Wells Rowe, 2009) och fortsätter under lång tid (Dockrell, 2009). Skrivaktiviteter genomsyrar undervisning och examination för praktiskt taget alla skolämnen, och elever ägnar en betydande del av skoldagarna åt aktiviteter kopplade till skrivande (McHale & Cermak, 1992). Samtidigt är det en stor utmaning för många barn att skriva. Skrivande är en komplex aktivitet som ställer krav på kognitiva, språkliga och motoriska förmågor, samt förmågan att koordinera dessa (Dockrell, 2009). Hur många barn som bedöms ha skrivsvårigheter beror på vilken definition som används. I en amerikansk kohortstudie av Katusic m.fl. (2009) uppskattades incidensen av ”writing disorder” till 6,9–14,7 %, beroende på vilka kriterier som användes. Motsvarande studier i svensk kontext saknas, men resultat från de nationella proven i svenska antyder att skrivande är svårt för många barn. Läsåret 2018/19 var andelen elever i årskurs 6 som inte nådde nivån för godkänt på delprovet ”skriva” 11,4 % på nationella proven i svenska, och 36,4 % i svenska som andraspråk (Skolverket, 2019). Att utveckla och utvärdera metoder som kan stötta skrivandet är därför ett angeläget forskningsområde.

Olika teoretiska modeller har tagits fram för att förklara de kognitiva processer som är inblandade i skrivande (se till exempel Hayes & Flower, 1980; Scardamalia & Bereiter, 1987). En modell som utvecklades av Berninger m.fl. (2002) för att förklara hur kognitiva processer är relaterade till skrivutveckling är ”The Simple View of Writing”, se figur 1. Modellen kan användas för att förklara hur svårigheter kan uppstå, vad dessa kan bero på och ge implikationer för hur skrivutveckling kan understödjas. Skrivprocessen förklaras utifrån en triangel med olika processer i respektive hörn.

I triangelns bas finns i ena hörnet transkription, vilket innefattar stavningsfärdigheter och motoriska färdigheter som att forma bokstäver eller tangentbordsskicklighet. I basens andra hörn finns exekutiva funktioner, vilket inkluderar uppmärksamhet, självreglering, att sätta mål, planera och revidera skrivandet. Transkription beskrivs som en lågnivåprocess och



Figur 1. Illustration av ”The Simple View of Writing” av Berninger m.fl. (2002) med författarnas översättning av de inblandade processerna.

exekutiva funktioner som en högnivåprocess. Båda dessa arbetar tillsammans för att understödja textgenerering, som återfinns i toppen av triangeln. Textgenerering är målet med skrivandet och här ingår idéer och begrepp som skribenten ska förmedla, inklusive de ord och meningar som används. Textgenereringen innefattar därmed både bakgrundskunskap, semantiska förmågor och syntaktiska förmågor (Dockrell, 2009). I mitten av triangeln återfinns arbetsminnet. Enligt modellen tävlar de olika delarna av skrivprocessen om de begränsade arbetsminnesresurserna. Om lågnivåfärdigheter som transkription inte är automatiserade, utan kräver mycket arbetsminnesresurser, finns mindre av dessa tillgängliga för processer på högre nivå (Berninger m.fl., 2002). Således kan stavningssvårigheter tänkas begränsa skribentens möjligheter att i skrivögonblicket generera idéer, översätta dessa till lämpliga ord och meningar, planera skrivandet och utföra revisioner av texten, och även hindra den långsiktiga utvecklingen av dessa förmågor. Argument för att transkriptionssvårigheter begränsar högnivåprocesser har anförts utifrån både teoretiskt perspektiv (McCutchen, 1996), psykologiska experiment (Bourdin & Fayol, 1994; Olive & Kellogg, 2002) och interventioner där man exempelvis tränat handskrivning och sett positiva effekter på högnivåprocesser som följd (Santangelo & Graham, 2016).

I föreliggande studie ligger fokus på barn med stavningssvårigheter. Förekomsten av stavningssvårigheter hos barn har undersökts både i populationsstudier och studier av kliniska grupper. Vid dyslexi ses stavningssvårigheter som ett centralt symptom som ofta kvarstår upp i åldrarna (Hatcher m.fl., 2002; Maughan m.fl., 2009). Stavningssvårigheter är också vanligt förekommande hos barn med språkstörning (Bishop & Clarkson, 2003) och hos barn med inlärningssvårigheter (Landerl & Moll, 2010). Populationsstudier har visat förekomst av stavningssvårigheter hos omkring 16 % av elever i låg- och mellanstadieålder (Landerl & Moll, 2010; Moll m.fl., 2014). Även om läsförmåga och stavningsförmåga ofta samvarierar (se till exempel Caravolas m.fl., 2001) kan stavningssvårigheter förekomma utan samtidiga lässvårigheter och vice versa (Moll & Landerl, 2009).

Förutom att stavningssvårigheter, enligt ”The Simple View of Writing”, tar resurser från textgenerering, har förekomst av stavfel i sig negativ påverkan på hur lärare bedömer textkvalitet (Graham m.fl., 2011; Marshall & Powers, 1969). Att underlätta transkriptionen för barn med stavningssvårigheter borde således förbättra textkvaliteten, genom att frigöra resurser och undvika negativa bedömningar till följd av stavfel. En metod som föreslagits för detta är att skriva med taligenkänning (MacArthur, 2009).

Att skriva med taligenkänning innebär att tala in text till ett program som överför talsignalen till skriven text (O’Shaughnessy, 2003). Till skillnad från traditionella skrivsätt, där texten transkriberas med penna eller med tangentbord, talar skribenten här in text som simultant dyker upp på skärmen. Mycket lite är känt om hur skrivprocessen påverkas av att skriva med taligenkänning. Eftersom verktyget i princip eliminerar stavningsaspekten av skrivandet, kan det tänkas gynna textkvaliteten för barn med stavningssvårigheter genom att arbetsminnesresurser kan omlokaliseras till andra delar av skrivandet. Å andra sidan är det möjligt att taligenkänning också kan innebära nya utmaningar. Skribenten behöver tala tydligt och uppmärksamma eventuella fel som verktyget gör. Taligenkänning skulle också, särskilt för ovana användare, kunna medföra en ökad belastning på exekutiva funktioner. Detta eftersom användaren kan sakna invanda strategier för att exempelvis planera skrivandet, uppmärksamma fel och revidera text med hjälp av verktyget. Den förmodade

resursvinsten som elimineringen av stavning innebär, skulle alltså kunna motverkas av en ökad exekutiv belastning.

Ett fåtal studier har undersökt om taligenkänning är ett bra hjälpmedel vid skrivsvårigheter. MacArthur och Cavalier (2004) fann att elever i gymnasieålder med skrivsvårigheter producerade bättre texter när de skrev med taligenkänning än för hand. Resultaten var dock tvetydiga. Quinlan (2004) fann inga skillnader i textkvalitet för barn i åldrarna 11–14. I en litteraturöversikt av Peterson-Karlan (2011) framkom att forskningen på taligenkänning vid skrivsvårigheter var för begränsad för att metoden skulle kunna klassificeras som evidensbaserad. Några ytterligare studier har gjorts sedan dess, med positiva resultat för bland annat andraspråksinlärare (Arcon m.fl., 2017) och universitetsstudenter (Nelson & Reynolds, 2015). Sammantaget finns dock stora skillnader avseende metod, vilket skrivsätt taligenkänning jämförts med, urvalskriterier och utfallsmått, och studierna har gett varierande resultat. Att avgöra huruvida taligenkänning kan underlätta vid skrivsvårigheter för barn i Sverige försvåras ytterligare av att studierna oftast genomförts i en engelskspråkig kontext med äldre taligenkänningsteknik vars korrekthet kan tänkas skilja sig mellan olika språk. Det finns alltså ett stort behov av att utvärdera hur skrivande med taligenkänning påverkar textkvalitet i en nutida svensk kontext, vilket föreliggande studie ämnar göra.

En relaterad fråga är huruvida texter skrivna med taligenkänning skiljer sig från texter skrivna för hand eller med tangentbord med hänsyn till andra egenskaper än bedömd textkvalitet. Skrivande med taligenkänning utgör ett slags mellanting mellan tal och skrift, i och med att skrivprocessen är muntlig medan slutprodukten är skriftlig. Det kan därför tänkas att dessa texter blir mer talspråkliga. För att mäta talspråkighet i texter kan måtten lexikal densitet och lexikal diversitet användas. Dessa utgör mått på andel ord från öppna ordklasser respektive andel unika ord i en text (Crossley, 2020). Johansson (2009) visade att lexikal diversitet och lexikal densitet är högre i skrift än i tal för barn utan svårigheter. Någon motsvarande skillnad mellan tangentbordsskrivna texter och texter skrivna med taligenkänning kunde dock inte påvisas av Kraft m.fl. (2019), som undersökte barn med stavningssvårigheter i åldrarna 10–13 år. En annan variabel som undersökts är förekomst av skrivfel i texterna. Taligenkänningsteknik minskar visserligen risken för stavfel, men medför samtidigt en risk för att andra typer av fel uppstår. Till exempel kan verktyget missa ord eller delar av ord, alternativt skriva ett annat ord än det som användaren avsett. Studier som jämfört andelen fel i texter skrivna med taligenkänning med handskrivna texter har visat att barn med skrivsvårigheter gör färre skrivfel med taligenkänning (MacArthur & Cavalier, 2004; Quinlan, 2004). Eftersom studierna inte skilt på olika typer av skrivfel, till exempel stavfel, taligenkänningsfel eller grammatiska fel, är det dock osäkert vad denna skillnad har för betydelse. Detta eftersom olika typer av fel skulle kunna påverka läsaren på olika sätt. Även textlängd har intresserat forskare, eftersom taligenkänningstekniken kan tänkas göra det möjligt för barn med skrivsvårigheter att producera längre texter. Detta har dock inte kunnat bekräftas entydigt i forskningen: Quinlan (2004) fann att barnen med skrivsvårigheter producerade längre texter med taligenkänning än för hand, medan MacArthur och Cavalier (2004) inte fann motsvarande skillnad. Inte heller Kraft m.fl. (2019), som dock jämförde texter skrivna med taligenkänning respektive med tangentbord, såg några skillnader mellan skrivsätten. Huruvida taligenkänning möjliggör ett mer avancerat vokabulär för barn med stavningssvårigheter, genom att stavningsaspekten elimineras, har undersökts i ett fåtal studier. Varken MacArthur och Cavalier (2004), som undersökte förekomst av lågfrekventa ord, eller Kraft m.fl. (2019), som undersökte

ordlängd, kunde påvisa en sådan effekt. Som framgår ovan är antalet studier på området få, resultaten delvis motstridiga, och endast enstaka variabler har undersökts. Hur skrivande med taligenkänning påverkar exempelvis syntaktisk komplexitet eller olika typer av fel, är till författarnas vetskap inte undersökt.

Om texter skrivna med taligenkänning faktiskt skiljer sig från tangentbordsskrivna texter avseende olika språkliga egenskaper, har detta någon betydelse för hur texternas kvalitet bedöms av läsaren? Många studier har undersökt vilka egenskaper i texter som har betydelse för hur textkvalitet bedöms. Sambanden mellan å ena sidan olika lingvistiska variabler och å andra sidan bedömningar av textkvalitet, sammanfattas i en översikt av Crossley (2020). I översikten framkommer att textlängd, andel avancerade ord ("lexical sophistication"), syntaktisk komplexitet och kohesion är de lingvistiska variabler som tenderar att uppvisa starkast samband med textkvalitet. Enligt författaren bör dock resultaten tolkas med viss försiktighet eftersom det finns stora skillnader i hur textkvalitet definierats och bedömts, vilka mått som använts för variablerna samt begränsningar i de automatiska analysverktyg som använts. I föreliggande studie kommer *textlängd*, *ordlängd*, *lexikal diversitet*, *lexikal densitet*, *syntaktisk komplexitet* samt *korrekthet* att analyseras. Variablerna är valda utifrån tidigare forskning om taligenkänning samt översikten av Crossley (2020) med anpassning till vilka analyser som är möjliga att göra automatiskt på svenskspråkiga texter.

Sammanfattningsvis är det alltså osäkert om taligenkänning är ett bra hjälpmedel vid stavningssvårigheter. Forskningen om hur textkvalitet och lingvistiska variabler påverkas är begränsad och det saknas studier om huruvida olika delar i skrivprocessen underlättas eller belastas mer vid skrivande med verktyget. I teorin borde barn med stavningssvårigheter och barn utan stavningssvårigheter påverkas på olika sätt av att skriva med verktyget, eftersom den förra gruppen får sina stavningssvårigheter underlättade medan den andra gruppen inte borde få någon liknande effekt. En kartläggning av hur skrivande med taligenkänning påverkar textkvalitet och lingvistiska variabler för båda grupperna, skulle kunna generera viktig information om vilka för- och nackdelar som finns med att använda taligenkänning som hjälpmedel vid stavningssvårigheter.

Syftet med denna studie är därför att undersöka om skrivande med taligenkänning påverkar textkvaliteten för barn med respektive utan stavningssvårigheter. Ett ytterligare syfte är att kartlägga eventuella lingvistiska skillnader mellan tangentbordsskrivna texter och texter skrivna med taligenkänning, samt att undersöka vilka variabler som påverkar hur textkvalitet bedöms.

### *Frågeställningar*

1. Hur påverkas textkvalitet i barns texter om de skrivs med taligenkänning jämfört med tangentbord, och skiljer sig dessa resultat för barn med respektive utan stavningssvårigheter?
2. Vilka likheter och skillnader avseende lingvistiska variabler finns mellan tangentbordsskrivna texter och texter skrivna med taligenkänning för barn med respektive utan stavningssvårigheter?
3. Vilka lingvistiska variabler i tangentbordsskrivna texter respektive texter skrivna med taligenkänning har ett samband med textkvalitet?

## Metod

Föreliggande studie är en del av det större forskningsprojektet ”Att skriva är silver och tala guld – eller?” (Dnr. MAW 2014.0122) som bedrivs vid Institutionen för svenska språket vid Göteborgs universitet. Syftet med projektet är att undersöka hur taligenkänning påverkar skrivandet för personer med läs- och skrivsvårigheter. I projektet har barn med respektive utan läs- och skrivsvårigheter gjort olika språkliga och kognitiva tester samt skrivit ett flertal texter med tangentbord och med taligenkänning. Projektet, inklusive föreliggande studie, är finansierat av Marcus och Amalia Wallenbergs minnesfond och har genomgått etikprövning (Dnr. 702–17). Barnen och deras vårdnadshavare har givit skriftligt samtycke till att medverka i projektet och samtliga analyser har gjorts på oidentifierad data.

I föreliggande studie användes texter insamlade i projektet samt delar av testresultaten. Nedan kommer först en beskrivning av rekryteringen av deltagarna och textinsamlingen som gjordes i det större projektet. Därefter beskrivs de textanalyser som gjordes i föreliggande studie.

### *Deltagare*

Femtioåtta barn mellan 10–13 år rekryterades till projektet genom bekvämlighetsurval. Deltagarna kom från åtta olika skolor i fyra olika kommuner i södra Sverige och Mellansverige. Skolornas socioekonomiska index varierade mellan 48–156, där 100 är det nationella genomsnittet och högre värden indikerar att en större andel elever på skolan riskerar att bli obehöriga till ett nationellt gymnasieprogram (Statens offentliga utredningar, 2017). Eleverna rekryterades via specialpedagoger och lärare som särskilt uppmuntrade elever med läs- och skrivsvårigheter att delta i projektet.

För att få delta i projektet skulle barnen gå i årskurs 3–6. Några barn hann dock börja årskurs 7 innan all data var insamlad. Dyslexidiagnos var inget inklusionskriterium och information om detta samlades heller inte in, med motiveringen att konsensus kring diagnossättning saknas och stavningssvårigheter även kan förekomma hos barn utan dyslexi. Däremot exkluderades barn med diagnostiserad utvecklingsstörning eller autismspektrumstörning. Flerspråkiga barn som gått hela sin skolgång i Sverige inkluderades i projektet. Ingen av deltagarna uppgav att de tidigare använt taligenkänning som skrivhjälpmedel.

Tabell 1

#### *Deltagarinformation.*

	Totalt	Stavningssvårigheter	Ej stavningssvårigheter
Antal	58 <sup>a</sup>	38	19
Flickor	25	17	8
Pojkar	33 <sup>a</sup>	21	11
Ålder (år) ( <i>M(±s)</i> )	12,0 (±1,0)	12,2 (±1,0)	11,7 (±0,9)

<sup>a</sup>Det saknas DLS-resultat för en deltagare. Deltagaren skrev en text som har inkluderats i korrelationsanalyserna. I gruppanalyserna har deltagaren exkluderats.

I föreliggande studie har deltagarna delats in i två grupper utifrån resultat på det normerade stavningstestet Rättstavning 2 ur DLS (Järpsten & Taube, 2010). Se tabell 1. I enlighet med Kraft m.fl. (2019) räknades deltagare med resultat lika med eller under stanine 3 till gruppen med stavningssvårigheter. Deltagare med resultat lika med eller högre än 4 räknades till gruppen utan stavningssvårigheter. Med utgångspunkt i ”The Simple View of Writing” (Berninger m.fl., 2002) bör man kunna förvänta sig att stavningssvårigheter inverkar negativt på textkvalitet. I och med att dessa svårigheter bör minska vid skrivande med taligenkänning bedömdes rättstavningstestet vara det tillgängliga mått som bäst skulle utskilja de deltagare som taligenkänningen hade störst potential att hjälpa, även om andra kognitiva och språkliga förmågor också kunde tänkas påverka utfallet.

### *Tillvägagångssätt*

*Textinsamling.* Texterna som användes i föreliggande studie samlades in mellan 2018–2019 av två logopedier i det ovannämnda forskningsprojektet. Deltagarna fick träffa en av de två logopederna enskilt under två tillfällen i de flesta fall. I några fall behövdes dock upp till fyra tillfällen på grund av bristande ork eller av schematekniska skäl. Under dessa tillfällen genomfördes först språkliga och kognitiva test och därefter textuppgifter. Se bilaga 1. Textuppgifterna utgick från filmklipp som illustrerade fusk respektive stöld, där uppgiften var att redogöra för hur en superhjälte skulle resonera kring situationen. Sammanlagt producerade deltagarna tre skrivna texter: en med tangentbord, en med taligenkänning och en tredje, så kallad kontrolltext, som var skriven antingen med tangentbord eller med taligenkänning. Deltagarna fick 30 minuter på sig att skriva varje text och hade inte tillgång till stavningskontroll. När de skrev med taligenkänning fick de använda tangentbordet, vilket somliga gjorde till exempel när taligenkänningen skrev ett annat ord än ordet de hade sagt. Efter textinsamlingen korrigerades texterna för uppenbara stavfel respektive taligenkänningsfel. Detta innebär att morfologiska och syntaktiska fel i texterna lämnades kvar. För detaljer om rättningen hänvisas till Kraft m.fl. (2019).

I föreliggande studie har de korrigerade texterna använts. Detta av flera anledningar. Dels för att läsarna inte skulle kunna avgöra om texterna var skrivna med tangentbord eller med taligenkänning, dels krävdes det att stav- och taligenkänningsfel var korrigerade för att kunna göra tillförlitliga lexikala analyser. Därutöver syftade studien till att undersöka hur övriga aspekter av skrivandet påverkas när stavningsaspekten försvinner. Hade texterna bedömts med stav- och taligenkänningsfel hade övriga aspekter riskerat att hamnat i skymundan av felen och påverkat bedömningarna i mindre utsträckning, vilket hade gjort det svårt att undersöka det studien syftade till att undersöka.

*Bortfall.* Eftersom varje deltagare skulle producera tre texter skulle sammanlagt 174 texter producerade med tangentbord och taligenkänning samlats in. En deltagare producerade dock fyra texter eftersom en text felaktigt troddes ha försvunnit. Det fanns också deltagare som inte skrev samtliga tre texter. Totalt var det 24 texter som föll bort. Fyra av dessa var av tekniska skäl. En deltagare dikterade en text på engelska som därför fick exkluderas. Fyra texter föll bort till följd av avhopp. Femton texter, samtliga kontrolltexter, föll bort på grund av tidsbrist. Arton av texterna som föll bort var skrivna av barn med stavningssvårigheter, varav 14 av dessa var kontrolltexter. Fyra av texterna som föll bort var skrivna av barn utan stavningssvårigheter, varav alla var kontrolltexter.



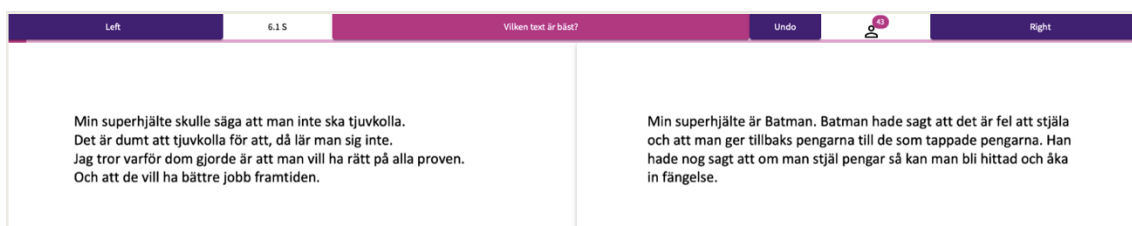
För två av texterna som föll bort saknades information om huruvida barnet hade stavningssvårigheter. Efter bortfall återstod sammanlagt 151 texter.

*Texter i föreliggande studie.* I föreliggande studie bedömdes och analyserades samtliga 151 texter som samlats i det större forskningsprojektet. Endast deltagarnas två första texter användes emellertid för statistiska analyser. Den tredje texten, kontrolltexten, användes enbart för en kontrollanalys som visade att ordningsföljd och vilket filmklipp deltagarna skrivit utifrån inte påverkade resultaten.

*Bedömning av textkvalitet.* Metoden Comparative Judgement (CJ) användes för att bedöma textkvalitet. I denna metod används parvisa jämförelser mellan texter där bedömare enskilt ska välja vilken av de två texterna som är bäst, se figur 2. Varje bedömare gör ett antal sådana jämförelser och baserat på resultaten från samtliga bedömningar räknar en matematisk modell fram ett kvalitetsvärde för respektive text (Pollitt, 2012). Detta kvalitetsvärde konverteras till ett så kallat ”scaled score”, som i föreliggande studie ligger mellan 0–100. Den text som utifrån den matematiska modellens beräkningar bedömts som sämst får ”scaled score” 0, den text som bedömts som bäst får ”scaled score” 100 och resterande texter får värden däremellan. Skalan är en kvotskala vilket bland annat innebär att flera texter kan ha samma ”scaled score”. Det innebär också att en dubbling i ”scaled score” innebär en dubbling i textkvalitet, till exempel har en text med ”scaled score” 50 bedömts som dubbelt så bra som en text med ”scaled score” 25. ”Scaled score” kommer fortsättningsvis benämnas som textkvalitet.

En av fördelarna som lyfts fram med CJ är hög reliabilitet (Pollitt, 2012; Steedle & Ferrara, 2016; Verhavert m.fl., 2019), vilket ofta är svårt att uppnå med kriteriebaserade bedömningsmetoder (se till exempel Skolinspektionen, 2011; Skolverket, 2009). Att metoden ger höga reliabilitetsvärden kan bero på att bedömarna inte behöver kalibreras gällande hur bra en text behöver vara för att nå en viss nivå, vilket är fallet i kriteriebaserade bedömningar, utan endast behöver vara eniga om vilken som är bäst (Pollitt, 2012). En annan fördel med att använda metoden i föreliggande studie var att textkvalitet inte på förhand behövde definieras och delas upp i mindre beståndsdelar, vilket hade riskerat att styra bedömningarna i en riktning som gynnade ett av skrivsätten mer än det andra. Konsekvensen av ett sådant förfarande skulle kunna bli att resultatet berodde på förhandsgivna kriterier snarare än faktisk textkvalitet. Detta eftersom det är möjligt att de två skrivsätten underlättar och försvårar olika delar av skrivandet.

Webbsidan Nomoremarking.com (2021) användes för att göra CJ-bedömningar av samtliga 151 texter. De två författarna till föreliggande studie, en av forskarna i



Figur 2. Parvis jämförelse med Comparative Judgement (publicerad med deltagarnas godkännande).

huvudprojektet samt en forskare med inriktning läs- och skrivsvårigheter vid förvärvat språkstörning deltog i bedömningarna. Bedömningarna gjordes enskilt. Ingen av bedömarna hade deltagit i textinsamlingen. Inför bedömningarna gjordes en pilotstudie där de fyra bedömarna gjorde CJ-bedömningar av texter som liknade texterna i föreliggande studie. Detta för att säkerställa att bedömarna var någorlunda samstämmiga. Därefter sammanställde författarna ett antal instruktioner för bedömningarna i föreliggande studie. Dessa inkluderade att bedömarna skulle använda sin expertis och göra en sammanvägd bedömning av textkvaliteten, vilket syftade till att bedömningarna skulle baseras på holistisk textkvalitet snarare än enskilda kriterier. För att inte eventuell inkonsekvens i rättningen av texterna skulle påverka bedömningarna, skulle bedömarna försöka bortse från eventuella stavfel eller taligenkänningsfel, utifall sådana skulle kvarstå trots rättningen. Bedömarna skulle också före varje bedömningsomgång läsa igenom instruktionen som utgått till deltagarna. Detta för att i bedömningarna beakta huruvida deltagarna följt skrivuppgiften, vilket dock inte skulle bedömas som ett kriterium överordnat andra aspekter av textkvalitet. Det rekommenderades att ta paus efter max 20 bedömningar och sprida ut bedömningarna över flera dagar för att undvika att trötthet skulle påverka bedömningarna. Tjugo parvisa jämförelser gjordes per text vilket resulterade i sammanlagt 3020 bedömningar. Antalet parvisa jämförelser per text bestämdes utifrån en metaanalys av Verhavert m.fl. (2019) som visade att i regel 19-20 jämförelser krävs för att uppnå en "Scale Separation Reliability" (SSR) på ,80. SSR anges i ett tal mellan 0 till 1 och används ofta som reliabilitetsmått i CJ. Måttet har visat sig motsvara både "split-half"-reliabilitet (Verhavert m.fl., 2018) och Cronbachs alfa (Pollitt, 2012). I föreliggande studie uppmättes SSR till ,92. I CJ erhålls även ett mått, "infit", på hur samstämmig respektive bedömare är med den sammanlagda skalan. Värdet varierar från 0 till oändlighet där ett lägre värde innebär högre samstämmighet (Weadon, 2018). Värdet över 1,2 anses som höga och det rekommenderas att man analyserar orsaken till detta för att eventuellt utesluta bedömaren. Ett högt "infit"-värde behöver inte nödvändigtvis vara negativt ur validitetssynpunkt. En avvikande bedömare kan tänkas tillföra aspekter av textkvalitet till bedömningarna som övriga bedömare missat, vilket i sådana fall kan bidra till högre validitet (van Daal m.fl., 2016). I föreliggande studie hamnade en bedömare över den angivna gränsen (1,4). Inga systematiska avvikelser hos denna bedömare kunde upptäckas vid den följande analysen. Eftersom resultaten endast påverkades marginellt om bedömaren uteslöts, behölls resultaten för samtliga bedömare.

*Analys av lingvistiska variabler.* Inför analysen av lingvistiska variabler (se tabell 2) gjordes enstaka korrigeringar av texterna manuellt av författarna i syfte att analysprogrammen skulle kunna göra korrekta och sinsemellan konsekventa analyser. Siffror skrevs ut med bokstäver, ord som satt samman med bindestreck skrevs samman och ord som satt samman med skiljetecken skrevs isär. Enligt detta korrigerades "100-dollarsedeln" till "hundraollarsedeln" och analyserades som ett ord, medan "t.ex." tillfogades blanksteg och analyserades som två ord. En av texterna utgjordes av till synes slumpmässigt valda ord utan inbördes samband eller relevans för uppgiften, och exkluderades därför från de lingvistiska analyserna.

Textlängd och andel långa ord, definierat som det totala antalet ord i texten respektive hur stor andel av dessa ord som var över sex bokstäver, mättes på Lix.se (2021). Lexikal diversitet, som mäts genom att analysera förekomsten av unika ord i relation till det totala antalet ord (Crossley, 2020), analyserades på Textinspector.com (2021). Verktuget

tillhandahåller flera mått på lexikal diversitet. Måttet som valdes i denna studie var ”Measure of textual lexical diversity” (MTLD), eftersom detta visat sig påverkas mindre av textlängd än alternativa mått (McCarthy & Jarvis, 2010). För texter kortare än 50 ord kan emellertid inte lexikal diversitet beräknas (McCarthy & Jarvis, 2010). Lexikal densitet definierades som antalet ord från öppna ordklasser i relation till det totala antalet ord. För att beräkna detta upprättades först en frekvenslista över samtliga ord i samtliga texter som ordklasstagades i Swegram (2021). Sedan korrigerades listan manuellt och orden kategoriserades som tillhörande antingen öppna eller stängda ordklasser. Som öppna ordklasser räknades verb, substantiv, adjektiv och adverb som avletts från adjektiv, i enlighet med till exempel Johansson (2009) och Kraft m.fl. (2019). Utifrån kategoriseringen upprättades en korpus för orden från öppna ordklasser, som sedan varje text kördes mot i AntWordProfiler (Anthony, 2012) för att beräkna hur stor andel av orden i texten som tillhörde öppna ordklasser.

Inför analyserna av syntaktisk komplexitet och korrekthet delades texterna upp i t-units och bisatser av författarna som analyserade hälften av texterna var. T-unit definieras som den minsta avslutningsbara syntaktiska enheten (Hunt, 1965) och kan bestå av en huvudsats plus eventuella bisatser som är underordnade denna (Crossley, 2020). Därmed kan en t-unit bestå av enbart en huvudsats, eller en huvudsats och flera underordnade bisatser, men inte flera huvudsatser eller enbart bisatser. Enligt Hunt (1965) är en fördel med att använda t-units som analysenhet att dessa bättre fångar den underliggande

Tabell 2

*Mått och tillvägagångssätt vid analys av lingvistiska variabler.*

<i>Lingvistisk variabel</i>	<i>Mått</i>	<i>Tillvägagångssätt</i>
Textlängd	Antal ord	Automatisk analys på Lix.se
Ordlängd	Andel långa ord (>6 bokstäver)	Automatisk analys på Lix.se
Lexikal diversitet	Measure of textual lexical diversity (MTLD)	Automatisk analys i Textinspector
Lexikal densitet	Andel ord från öppna ordklasser	Upprättande av frekvenslista (korpus) och ordklasstagning i Swegram Manuell korrigering av listan Matchning av texter mot korpus i AntWordProfiler
Syntaktisk komplexitet	Antal bisatser per t-unit	Manuell klassificering av t-units
Andel morfologiskt-syntaktiskt korrekta t-units	Antalet morfologiskt-syntaktiskt korrekta t-units dividerat med det totala antalet t-units	Manuell bedömning av morfologiska och syntaktiska fel
Andel fullständiga t-units	Antalet t-units som inte saknade ord eller satsled dividerat med det totala antalet t-units	Manuell bedömning av fullständighet
Andel korrekt avslutade t-units	Antalet t-units som avslutats korrekt med lämpligt skiljetecken eller konjunktion dividerat med det totala antalet t-units	Manuell bedömning av korrekt avslutning

syntaktiska strukturen än vad som är fallet för meningar. Detta eftersom

meningsuppdelning och meningslängd starkt påverkas av skribentens förmåga eller preferenser gällande interpunktion (Hunt, 1965). Till skillnad från meningar, som är tydligt grafiskt avgränsade genom stor bokstav och punkt, kräver emellertid t-unit-klassifikation en djupare analys med på förhand definierade ställningstaganden till hur exempelvis anföranden, interjektioner och ofullständiga huvud- och bisatser ska behandlas. Eftersom texterna visade sig innehålla många ofullständiga satser bedömdes det angeläget att formulera tydliga kriterier för klassificeringen. Ett antal kriterier fastställdes inför analysen. När kriterierna visade sig vara ofullständiga för att hantera oklarheter, diskuterades fallen för att nå konsensus. Därefter reviderades kriterierna och klassificeringen gjordes om för berörda texter. Den slutliga klassificeringen hade några tillägg och undantag till definitionen given av Hunt (1965). Samordnade verbfraser med underförstått subjekt ("Sen kom en tjej och hittade den på marken och plockade upp den") klassificerades som en t-unit, i enlighet med Bishop och Clarkson (2003). Några undantag gjordes till det annars gällande kriteriet att en t-unit skulle innehålla en fullständig huvudsats. Fragment, interjektioner och andra ofullständiga satser räknades som självständiga t-units om de inte uppenbart tillhörde föregående eller efterföljande t-units, bedömt utifrån den syntaktiska och semantiska kontexten. Till exempel räknades "En klassrum och barnen gör prov" som två t-units, eftersom samhörigheten mellan leden bedömdes svag. Interjektionerna i "Ok hejdå förlåt svarar tjejen" bedömdes dock tillhöra samma t-unit eftersom de ansågs som objekt till verbet "svarar". På samma sätt räknades "Då tänker han som tog upp saken jag ska nog ge tillbaka den" som en t-unit då huvudsatsen "jag ska nog ge tillbaka den" räknades som objekt till verbet "tänker". Vid liknande anföranden då det följde flera huvudsatser räknades dock endast den första tillhöra t-uniten. Enligt denna princip räknades till exempel "då kommer dom säga vi hann inte eller vi orkade inte" alltså som två t-units. Bisatser med underförstådd huvudsats, till exempel "[Superhjälten skulle tycka] Att man skulle plocka upp den och lämna tillbaka den", räknades som en t-unit.

Bisatsindelningen utgick från den definition som ges i Svenska Akademiens grammatik (Teleman, 1999). Denna innefattar bland annat att bisatser är syntaktiskt underordnade, inleds med subjunktion eller satsbas och att de semantiskt inte utgör självständiga språkhandlingar. För att klassificeras som bisats krävdes att det fanns en subjunktion. Detta gällde dock inte i fall då bisatsinledaren ansågs vara optionell, som i "Han tycker [att] det måste bli striktare regler" och i konditionala bisatser av typen "Kan man ingenting får man inget jobb". Vidare krävdes att det fanns ett finit verb. Undantag för detta gjordes dock i enstaka fall då kontexten tydligt pekade mot att det var en bisats men där verbet felaktigt skrivits i infinitiv form. Infinitivfraser med möjlig bisatsfunktion klassificerades inte som bisatser eftersom denna gränsdragning bedömdes vara för svår att bedöma konsekvent. För att mäta syntaktisk komplexitet räknades antalet bisatser per t-unit ut för respektive text. Även t-unit-längd beräknades men uteslöts då det visade sig korrelera starkt med antal bisatser per t-unit.

Varje texts samtliga t-units analyserades avseende korrekthet utifrån tre aspekter: morfologi/syntax, fullständighet samt förekomst av avslutande interpunktion eller konjunktion. Aspekterna valdes utifrån att många texter verkade innehålla fel inom dessa områden. *Andel morfologiskt-syntaktiskt korrekta t-units* beräknades genom att för varje text dela antalet morfologiskt-syntaktiskt korrekta t-units med det totala antalet t-units. En t-unit bedömdes som morfologiskt-syntaktiskt inkorrekt om det förekom exempelvis

felaktig genus-, numerus eller speciesböjning i ord eller fraser, uppenbart inkongruent tempusböjning av verb inom samma t-unit, eller ordföljdsfel. Smärre fel som bedömdes som konventionaliserade i talspråk, till exempel "denna situationen", räknades inte som fel. Felaktiga konstruktioner som syntaktiskt fungerar som enstaka ord/satsled, till exempel "fuska av [+objekt]" och "göra fel av sig" bedömdes vara lexikala fel och räknades därför inte heller. *Andel fullständiga t-units* utgjorde en underkategori till morfologiskt-syntaktiskt korrekta t-units. Antalet fullständiga t-units delades med det totala antalet t-units. En t-unit bedömdes som ofullständig i de fall då ett morfologiskt-syntaktiskt fel kunde avhjälpas genom att lägga till ett eller ett fåtal ord, utan att ta bort något ord, till exempel "Då var det en flicka bakom henne som tog upp pengarna [och] behöll dem för sig själv". I fall som antingen kunde tolkas som bortfall av ord eller som något annat morfologiskt/syntaktiskt fel, eftersträvades den enklaste tolkningen. *Andel korrekt avslutade t-units* beräknades genom att dela antalet t-units som avslutades genom antingen lämpligt skiljetecken eller konjunktion, med det totala antalet t-units. Om en t-unit avslutades med kommatecken betraktades detta som satsradning vilket räknades som inkorrekt avslut. Titel var undantagen kravet på avslutande skiljetecken eller konjunktion.

Klassificeringen av t-units, bisatser samt korrekthetsvariablerna beskrivna ovan gjordes av de två författarna, som genomförde hälften av klassificeringarna var. Särskilt svåra fall diskuterades tills konsensus uppnåddes. För att analysera interbedömarreliabilitet valdes på förhand 20 texter ut som enskilt skulle bedömas av båda författarna. Dessa texter var jämnt fördelade mellan skrivsätten och författarna var på förhand ovetande om texternas karaktär. Korrelationerna, uppmätta med Pearsons korrelationskoefficient, var för klassificering av t-units ,998, för bisatser ,998, för andel morfologiskt-syntaktiskt korrekta t-units ,957, för andel fullständiga t-units ,893 samt för andel korrekt avslutade t-units 1,000.

### *Statistisk analys*

De statistiska analyserna gjordes i SPSS version 27 och signifikansnivån sattes till  $p < ,05$ . För att undersöka hur textkvalitet påverkades av skrivsätt gjordes en tvåvägs ANOVA. Denna användes även för att analysera interaktionseffekter, alltså om det fanns någon skillnad i hur gruppernas textkvalitet påverkades av skrivsätt.

För att undersöka hur respektive deltagare påverkades av de två skrivsätten togs det för varje deltagare fram ett diskrepansvärde. Textkvaliteten för texten skriven med taligenkänning subtraherades med textkvaliteten för den tangentbordsskrivna texten. Detta resulterade i ett diskrepansvärde för respektive deltagare som antingen var positivt eller negativt. Ett positivt värde indikerar att deltagarens fick högst textkvalitet när den skrev med taligenkänning. Ett negativt värde indikerar att deltagaren fick högst textkvalitet när den skrev med tangentbord. Om en deltagare till exempel fick ett textkvalitetsvärde på 50 när de skrev med taligenkänning och 38 med tangentbord, resulterade det i ett positivt diskrepansvärde på 12.

Då det observerades att en större andel i gruppen med stavningssvårigheter än i gruppen utan stavningssvårigheter bedömdes få högre textkvalitet när de skrev med taligenkänning gjordes ett chi-2 test för att analysera om skillnaden var signifikant. Inför

chi-2 testet säkerställdes det att den uppmätta skillnaden inte var ett resultat av slumpen, genom att ta hänsyn till standardfelet för texternas respektive textkvalitetspoäng. I CJ anger standardfelet ett intervall som det faktiska värdet kan ligga inom. För en text med textkvalitet 36 och standardfel 4 kan alltså det faktiska värdet ligga mellan 32 och 40. Om diskrepansen i textkvalitet mellan texten skriven med tangentbord och texten skriven med taligenkänning var större än summan av de båda texternas standardfel, räknades det som en faktisk skillnad mellan de två texterna. Om diskrepansen var mindre än summan av texternas standardfel räknades det inte som en faktisk skillnad.

Icke-parametriska tester valdes för att analysera de lingvistiska variablerna. Detta eftersom variablerna inte var normalfördelade. ”Wilcoxon signed-rank test” användes för att se om det fanns en signifikant skillnad mellan tangentbordsskrivna texter och texter skrivna med taligenkänning för de båda grupperna gällande de lingvistiska variablerna. Holm-Bonferroni-metoden användes för att korrigera för multipla jämförelser. Liksom för textkvalitet, beräknades också ett diskrepansvärde för respektive deltagare. Detta gjordes för samtliga lingvistiska variabler. Spearmans rangkorrelation användes för att undersöka hur de lingvistiska variablerna korrelerade med textkvalitet.  $r_s < ,30$  tolkades som en försumbar korrelation,  $r_s = ,30- ,50$  tolkades som en svag korrelation,  $r_s = ,50- ,70$  tolkades som en måttlig korrelation,  $r_s = ,70- ,90$  tolkades som en stark korrelation och  $r_s > ,90$  tolkades som en mycket stark korrelation (Mukaka, 2012).

## Resultat

### *Skillnad i textkvalitet beroende på grupp och skrivsätt*

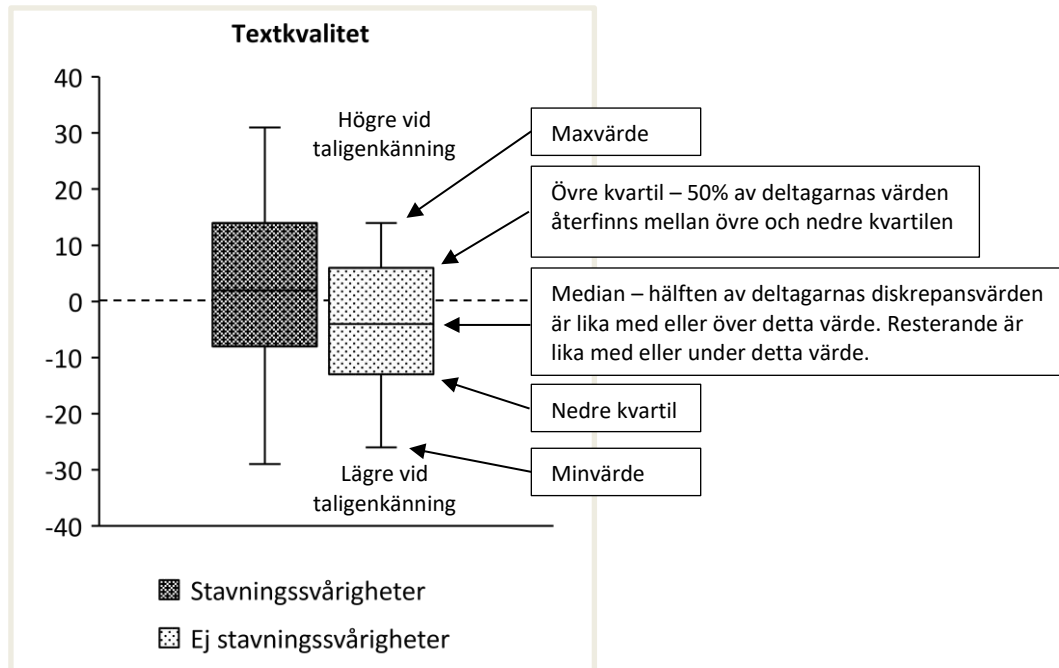
ANOVA-analysen av gruppernas medelvärden visade ingen signifikant skillnad i textkvalitet beroende på skrivsätt,  $F(1, 1) = ,54, p = ,46$ . Varken för gruppen med stavningssvårigheter eller för gruppen utan stavningssvårigheter påverkades alltså textkvalitet av skrivsätt. Det fanns inte heller någon signifikant interaktionseffekt, vilket innebär att grupperna inte påverkades olika av skrivsätt,  $F(1, 1) = 2,50, p = ,10$ . Däremot visade chi-2 testet en signifikant skillnad mellan grupperna gällande andelen barn som fick högre textkvalitet när de skrev med taligenkänning,  $\chi^2(1, 52) = 3,89, p < ,05$ . I gruppen med stavningssvårigheter fick 14 av 33 deltagare högre bedömd textkvalitet när de skrev med taligenkänning. För gruppen utan stavningssvårigheter var motsvarande andel 3 av 19. Generellt producerade gruppen utan stavningssvårigheter texter med högre textkvalitet än gruppen med stavningssvårigheter,  $F(1, 1) = 7,90, p < ,05$ .

### *Tabell 3*

Textkvalitet för de två grupperna vid de olika skrivsätten.

	Textkvalitet					
	Tangentbord			Taligenkänning		
	<i>N</i>	<i>M</i>	<i>s</i>	<i>N</i>	<i>M</i>	<i>s</i>
Stavningssvårigheter	33	51,6	13,9	33	53,2	11,2
Ej stavningssvårigheter	19	65,0	15,9	19	60,6	18,4

*Notering.* *M* = medelvärde, *s* = standardavvikelse



Figur 3. Hur deltagarnas textkvalitet påverkades av skrivsätt.

Tabell 3 visar den deskriptiva statistiken för de båda gruppernas textkvalitet med respektive skrivsätt. Figur 3 illustrerar diskrepansvärden, det vill säga hur textkvaliteten påverkades av skrivsätt för respektive deltagare. Ett positivt diskrepansvärde, det vill säga ett värde över nollstrecket, indikerar att deltagaren fick högre textkvalitet när den skrev med taligenkänning. Ett negativt diskrepansvärde, det vill säga ett värde under nollstrecket, indikerar att deltagaren fick lägre textkvalitet när den skrev med taligenkänning. Det fanns en stor variation i hur deltagarna i de två grupperna påverkades av skrivsätt. I gruppen med stavningssvårigheter fanns en större spridning i hur deltagarnas textkvalitet påverkades, jämför avstånden mellan max- och minvärde i de båda grupperna. För gruppen med stavningssvårigheter ökade textkvaliteten för den deltagare vars textkvalitet gynnades mest av att skriva med taligenkänning lika mycket som textkvaliteten minskade för den deltagare vars textkvalitet missgynnades mest av att skriva med taligenkänning, se max- och minvärde. Detta ses inte för gruppen utan svårigheter, där den största negativa diskrepansen är betydligt större än den största positiva, se max- och minvärde.

#### *Skillnad mellan skrivsätten för de olika lingvistiska variablerna*

Inte heller för de lingvistiska variablerna fanns några signifikanta skillnader mellan skrivsätten för någon av grupperna, efter korrektion med Holm-Bonferroni-metoden. Se tabell 4. Däremot fanns det, liksom för textkvalitet, en stor individuell variation i hur de lingvistiska variablerna påverkades av skrivsätt. Se figur 4 som visar diskrepansen mellan deltagarnas texter skrivna med taligenkänning och tangentbordsskrivna texter, alltså ökningen eller minskningen av variablerna när deltagarna skrev med taligenkänning. För ordlängd, mätt i andel långa ord, lexikal diversitet, mätt i MTLD, lexikal densitet, mätt i andel ord från öppna ordklasser, samt korrekthetsmåttet fanns tendenser till större spridning i gruppen med stavningssvårigheter. Huruvida denna var signifikant undersöktes inte.

Tabell 4

*Skillnader mellan skrivsätten gällande lingvistiska variabler.*

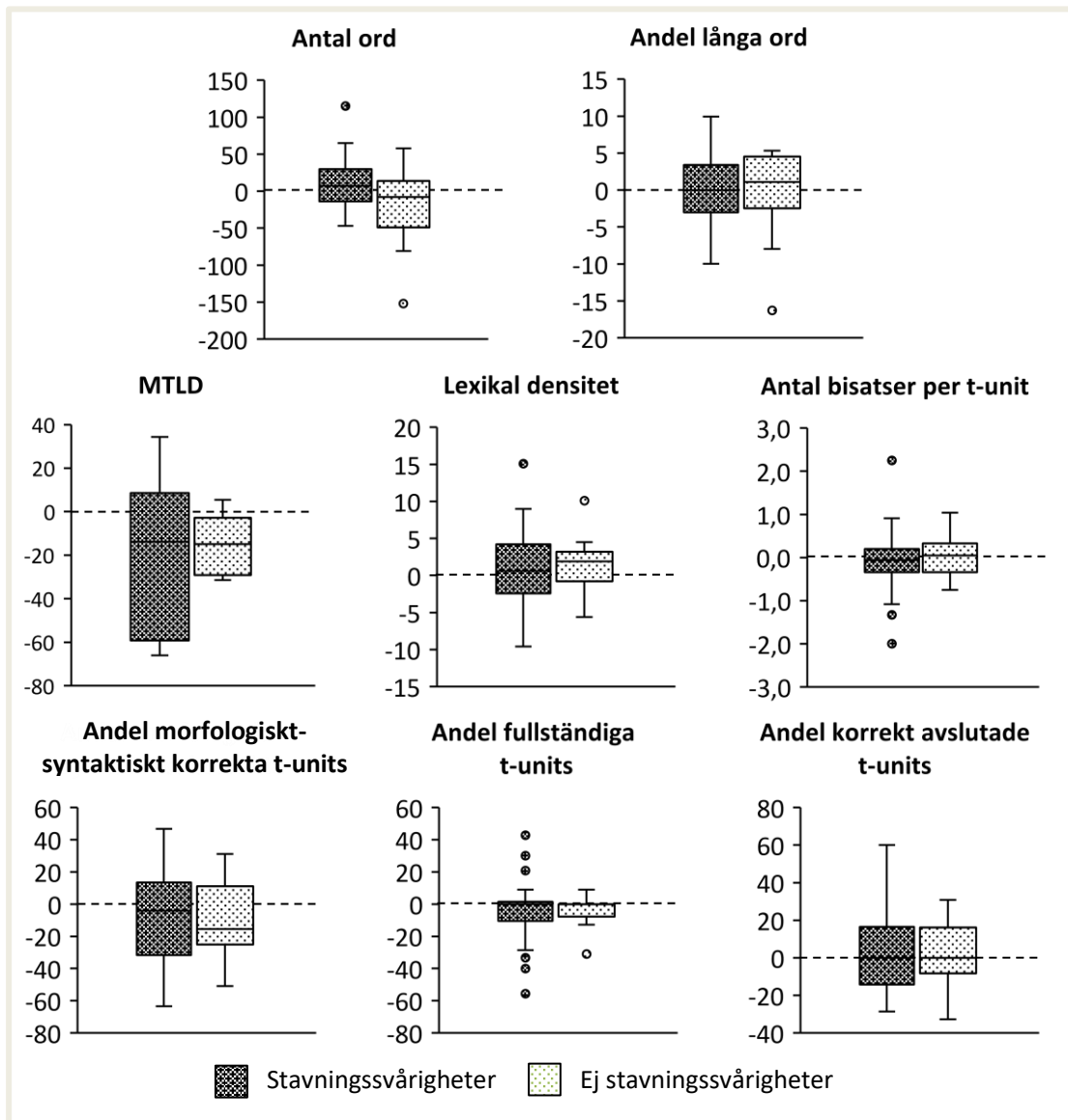
Lingvistisk variabel		Skrivsätt						Z	Sig.
		Tangentbord			Taligenkänning				
		N	M(s)	Med(min-max)	M	Med(min-max)			
Antal ord	Stavningsvårigheter	33	75,2(37,0)	76(15–173)	85,4(41,9)	74,4(20–192)	360,5	,15	
	Ej stavningssvårigheter	19	117(63,9)	102(34–245)	99,7(52,7)	102(30–212)	57	,22	
Andel långa ord	Stavningsvårigheter	33	11,4(4,3)	11,0(3,0–22,2)	11,6(5,1)	10,7(2,9–23,0)	269,5	,92	
	Ej stavningssvårigheter	19	12,6(4,6)	10,7(6,7–24,9)	12,4(3,4)	12,9(5,8–18,8)	91	,81	
MTLD	Stavningsvårigheter	21 <sup>a</sup>	50,2(14,0)	48,9(29,0–72,6)	44,4(18,2)	38,5(21,7–80,6)	78	,19	
	Ej stavningssvårigheter	15 <sup>a</sup>	63,5(20,8)	62,1(39,1–104,0)	50,3(14,0)	48,3(32,1–77,6)	38	,21	
Lexikal densitet	Stavningsvårigheter	33	44,2(4,9)	43,3(34,9–61,9)	45,4(4,0)	44,8(35,2–57,1)	332	,36	
	Ej stavningssvårigheter	19	45,7(4,5)	45,6(35,7–52,3)	47,3(3,4)	46,8(41,1–53,8)	142,5	,06	
Antal bisatser per t-unit	Stavningsvårigheter	33	1,0(,6)	,8(,3–3,0)	,9(,5)	,8(,2–2,8)	211,5	,47	
	Ej stavningssvårigheter	19	,9(,5)	,9(,3–2,0)	,9(,5)	,8(,3–1,7)	82	,80	
Andel morfologiskt-syntaktiskt korrekta t-units	Stavningsvårigheter	33	82,2(20,7)	85,7(33,0–100,0)	74,5(21,8)	75,0(22,2–100,0)	125,5	,13	
	Ej stavningssvårigheter	19	88,0(15,5)	92,3(50,0–100,0)	79,6(16,4)	78,9(40,0–100,0)	45	,14	
Andel fullständiga t-units	Stavningsvårigheter	33	92,7(13,7)	100,0(50,0–100,0)	89,5(14,7)	100,0(44,0–100,0)	58	,38	
	Ej stavningssvårigheter	19	98,2(3,3)	100,0(90,9–100,0)	94,2(9,8)	100,0(60,0–100,0)	5	,04 <sup>b</sup>	
Andel korrekt avslutade t-units	Stavningsvårigheter	33	85,2(19,9)	100,0(33,3–100,0)	88,9(12,5)	88,9(46,2–100,0)	197	,59	
	Ej stavningssvårigheter	19	87,2(15,3)	93,8(57,1–100,0)	88,4(15,6)	91,7(40,0–100,0)	43,5	,72	

*Notering.* N = antal deltagare, M = medelvärde, s = standardavvikelse, Med = median.

<sup>a</sup> Vid beräkning av MTLD krävs texter på minst 50 ord. Därför föll texter av 12 deltagare med stavningssvårigheter och fyra deltagare utan stavningssvårigheter bort.

<sup>b</sup> Efter korrektion med Holm-Bonferroni-metoden var värdet inte längre signifikant.





Figur 4. Hur skrivsätt påverkade de olika lingvistiska måtten för respektive deltagare.

#### *Samband mellan textkvalitet och de lingvistiska variablerna*

Det fanns en måttlig positiv korrelation mellan textkvalitet och textlängd, mätt i antal ord, för tangentbordsskrivna texter,  $r_s = ,62, p < ,05$ , och för texter skrivna med taligenkänning,  $r_s = ,69, p < ,05$ . Det fanns en svag positiv korrelation mellan textkvalitet och lexikal diversitet, mätt i MTLD, för tangentbordsskrivna texter,  $r_s = ,35, p < ,05$ , och för texter skrivna med taligenkänning,  $r_s = ,33, p < ,05$ . Det fanns även en svag positiv korrelation mellan textkvalitet och syntaktisk komplexitet, mätt i antal bisatser per t-unit, för tangentbordsskrivna texter,  $r_s = ,32, p < ,05$ , och för texter skrivna med taligenkänning,  $r_s = ,31, p < ,05$ . Därutöver fanns det för tangentbordsskrivna texter en försumbar positiv korrelation mellan textkvalitet och andel korrekt avslutade t-units,  $r_s = ,29, p < ,05$ . Detta resultat återfanns inte för texter skrivna med taligenkänning.

För ordlängd, mätt i andel långa ord, lexikal densitet, mätt i andel ord från öppna ordklasser, andel morfologiskt-syntaktiskt korrekta t-units samt andel fullständiga t-units fanns inga signifikanta korrelationer med textkvalitet för något av skrivsätten. Andel långa ord närmade sig dock signifikans,  $r_s = ,25$ ,  $p = ,06$ , för texter skrivna med taligenkänning. Se tabell 5.

Tabell 5

*Spearman's rangkorrelationer mellan textkvalitet och de lingvistiska variablerna*

Lingvistisk variabel	Textkvalitet	
	Tangentbord ( $N = 55^a$ )	Taligenkänning ( $N = 55^a$ )
Antal ord	,62*	,69*
Andel långa ord	,12	,25
MTLD	,35*	,33*
Lexikal densitet	,12	,08
Antal bisatser per t-unit	,32*	,31*
Andel morfologiskt-syntaktiskt korrekta t-units	,08	,21
Andel fullständiga t-units	,10	-,04
Andel korrekt avslutade t-units	,29*	,00

\*  $p < ,05$

<sup>a</sup> Vid beräkning av MTLD krävs texter på minst 50 ord. Därför kunde 12 texter från respektive skrivsätt inte analyseras.

## Diskussion

Syftet med studien var att undersöka hur skrivande med taligenkänning påverkar textkvaliteten för barn med respektive utan stavningssvårigheter. Resultatet visade att det inte fanns någon signifikant skillnad i textkvalitet beroende på skrivsätt för någon av grupperna. Detta resultat var förvånande. Utifrån ”The Simple View of Writing” (Berninger m.fl., 2002) antogs att skrivandet skulle underlättas för gruppen med stavningssvårigheter när taligenkänning användes, eftersom arbetsminnesresurser borde frigöras när stavningsaspekten eliminerades. Detta borde inte gälla i samma utsträckning för gruppen utan stavningssvårigheter vars transkription borde vara automatiserad i högre grad. Det antogs att det kunde finnas utmaningar med att skriva med taligenkänning, till exempel med att planera och revidera text, som kunde utgöra svårigheter för båda grupperna. Att grupperna inte skilde sig åt tyder på att deltagarna med stavningssvårigheter inte blev hjälpta i den utsträckningen som det förväntades, och att deltagarna utan svårigheter inte blev negativt påverkade av att skriva med ett nytt skrivverktyg. Resultatet överensstämmer med Quinlan (2004) som fann att skrivande med taligenkänning inte påverkade textkvalitet vare sig för barn med eller för barn utan stavningssvårigheter. Resultatet stämmer också delvis överens med MacArthur och Cavalier (2004), som visserligen fann en tendens till att taligenkänning gav förbättrad textkvalitet för gruppen med stavningssvårigheter, men inte fann någon signifikant interaktionseffekt. Sammantaget antyder detta att taligenkänning inte per automatik underlättar skrivande för barn med stavningssvårigheter. I föreliggande studie fanns emellertid en stor individuell variation i hur textkvaliteten påverkades. I båda grupperna fanns barn vars

textkvalitet förbättrades markant och barn vars textkvalitet försämrades markant av att skriva med taligenkänning. I gruppen med stavningssvårigheter var andelen barn vars textkvalitet förbättrades av taligenkänning signifikant större än i gruppen utan svårigheter. Detta antyder att det för en del barn med stavningssvårigheter ändå kan finnas en positiv effekt av att eliminera stavningen, även om detta inte gäller alla i denna grupp.

Utöver att undersöka hur skrivande med taligenkänning påverkar textkvalitet, syftade studien till att kartlägga eventuella lingvistiska skillnader mellan skrivsätten, samt att undersöka vilka variabler som påverkar hur textkvalitet bedöms. Inga skillnader mellan skrivsätten fanns för några av de lingvistiska variablerna som undersöktes. Eftersom skrivande med taligenkänning utgör ett slags mellanting mellan tal och skrift, förmodades dessa texter bli mer talspråkliga i form av lägre lexikal diversitet och lägre lexikal densitet. Några sådana tendenser återfanns alltså inte i resultatet. Detta överensstämmer med Kraft m.fl. (2019) som inte heller fann några skillnader mellan måtten för de två skrivsätten. Att dessa resultat överensstämde var väntat eftersom undersökningsgrupp och metod delvis överlappade mellan studierna. Resultaten antyder att barn klarar av att anpassa textproduktionen till skriftspråkliga konventioner i samma utsträckning för båda skrivsätten. Inte heller syntaktisk komplexitet påverkades av skrivsätt, vilket även Quinlan (2004) tidigare visat. Även detta kan tolkas som att texter producerade med taligenkänning har liknande karaktär som tangentbordsskrivna texter. Ett annat antagande, som inte bekräftades i resultaten, var att barnen med stavningssvårigheter skulle kunna producera längre texter med taligenkänning eftersom deras produktion inte längre skulle begränsas av deras stavningssvårigheter. Att textlängden inte påverkades av skrivsätt överensstämmer med Kraft m.fl. (2019) och MacArthur och Cavalier (2004), men inte med Quinlan (2004). En förklaring kan, liksom för den uteblivna effekten på textkvalitet, vara att verktyget istället belastade andra delar av skrivprocessen vilket motverkade effekten av att stavningen underlättades. Att de olika korrekthetsvariablerna inte påverkades av skrivsätt var oväntat. Taligenkänningsteknik förväntades utgöra större risk för särskilda typer av fel, såsom bortfall av ord till följd av att verktyget ibland gör felaktiga analyser av talsignalen. Detta bekräftades inte av resultaten. Det verkar vara lika lätt alternativt svårt att undvika sådana fel med båda skrivsätten. Värt att notera var dock att det, liksom för textkvalitet, fanns en stor individuell variation i hur i princip samtliga variabler påverkades av skrivsätt. Även om texterna på gruppnivå inte skilde sig åt i karaktär, kan det inte uteslutas att textkaraktär kan påverkas av skrivsätt för vissa barn.

Slutligen undersöktes vilka lingvistiska variabler som påverkade textkvalitetsbedömningarna. Inga av variablerna uppvisade starka samband med textkvalitet. Det starkaste sambandet återfanns mellan textlängd och textkvalitet, där en måttlig korrelation uppmättes för båda skrivsätten. I övrigt fanns enbart svaga eller försumbara korrelationer, och fyra av åtta variabler uppvisade inga samband alls med textkvalitet. Resultaten överstämde delvis med Crossley (2020) med undantag för andel långa ord som inte uppvisade något samband med textkvalitet i föreliggande studie. En förklaring till att så många av variablerna inte uppvisade samband med textkvalitet kan vara att sambanden för vissa variabler inte är linjära. Detta har föreslagits för lexikal diversitet (Nyström, 2000) och kan även tänkas gälla andra variabler, även om detta till författarnas kännedom inte undersökts. En ökning av en variabel kan tänkas bidra till ökad textkvalitet upp till en viss nivå, men en fortsatt ökning efter denna nivå kan leda till att textkvaliteten minskar. För exempelvis syntaktisk komplexitet, är det troligtvis så

att en viss komplexitet är gynnsam för bedömd textkvalitet men att det finns en gräns vid vilken ökad komplexitet gör texten svårförståelig. Sammantaget antyder resultatet att textkvalitet är komplext och svårt att förklaras utifrån enskilda lingvistiska variabler, åtminstone inte de som analyserades i föreliggande studie. Möjligtvis hade en analys av olika kombinationer av lingvistiska variabler, istället för variabler i isolation, bättre kunnat förklara bedömd textkvalitet. En sådan analys rymdes emellertid inte inom ramen för denna studie.

Sammanfattningsvis visade resultaten att taligenkänning inte per automatik underlättar skrivande för barn med stavningssvårigheter och att det för båda grupperna fanns en stor individuell variation i hur texterna påverkades av att skriva med taligenkänning. Detta antyder att det finns andra faktorer än stavningssvårigheter som har betydelse för huruvida taligenkänning underlättar skrivande. En sådan är träning och erfarenhet av att använda verktyget. På samma sätt som skrivutveckling med traditionella skrivsätt baseras på många års träning (Dockrell, 2009) kan också skrivande med taligenkänning förväntas utvecklas med ökad erfarenhet av skrivsättet. De olika processerna i "The Simple View of Writing" är delvis annorlunda vid skrivande med taligenkänning. Transkriptionen sker annorlunda, och processer som planering och revidering skiljer sig sannolikt från andra skrivsätt (Leijten, 2007). Att skriva med taligenkänning ställer därför troligen delvis annorlunda krav på skribenten. För att skrivandet ska underlättas krävs sannolikt en kombination av erfarenhet och explicit träning i hur verktyget ska användas. Exempel på detta kan vara instruktioner i hur man ökar precision i transkriptionen och undervisning i strategier för att planera, strukturera och revidera text skriven med taligenkänning.

Högnivåfärdigheter, som lexikala och syntaktiska förmågor, mottagaranpassning och genreanpassning, vilka taligenkänning förmodas frigöra resurser för, behöver också beaktas. Många barn med stavningssvårigheter kan, på grund av sina svårigheter, ha blivit begränsade i utvecklingen av dessa färdigheter (Berninger m.fl., 2002). Detta kan bero på att stavningen tar så mycket arbetsminnesresurser i anspråk att resurser för att uppmärksamma och utveckla andra skrivfärdigheter saknas. Stavningssvårigheter samexisterar också relativt ofta med andra utvecklingsrelaterade svårigheter, som exempelvis uppmärksamhetssvårigheter (Brimo m.fl., 2021), vilket kan utgöra ytterligare hinder för att utveckla skrivfärdigheter på högre nivå. Således kan det inte förväntas att högnivåfärdigheter automatiskt kommer tillgängliggöras för barn med stavningssvårigheter, enbart för att stavningsaspekten försvinner. I linje med detta finns studier som visat att enbart träning av lågnivåfärdigheter inte leder till högre textkvalitet (Berninger m.fl., 2002; Graham m.fl., 2018). Sannolikt krävs även insatser inriktade på högnivåfärdigheter. Detta styrks av Berninger m.fl. (2002) och Quinlan (2004) som båda fann att textkvaliteten för barn med skrivsvårigheter förbättrades mer om underlättande av lågnivåprocesser kombinerades med undervisning i högnivåfärdigheter. Även om studierna skiljer sig åt avseende vilka lågnivå- respektive högnivåfärdigheter som undersökts, tyder resultaten på att interventioner inriktade på flera olika skrivfärdigheter är att föredra. Bäst utfall av taligenkänningsintervention skulle sannolikt uppnås om instruktioner och träning anpassades både efter barnets förmågor och de specifika villkor som skrivsättet innebär.

## *Metoddiskussion*

Skrivande, textkvalitet och skrivsvårigheter är komplexa områden vilket har inneburit vissa metodologiska utmaningar. Det saknas till exempel konsensus kring hur textkvalitet ska definieras, mätas och bedömas i såväl forskning som praktik (Van Steendam m.fl., 2012). Studier har visat låg överensstämmelse mellan enskilda bedömare då man använt kriteriebaserade metoder (se till exempel Skolinspektionen, 2011; Skolverket, 2009). Detta var en av orsakerna till att CJ användes som metod i föreliggande studie. Metoden är dock relativt ny i svensk kontext och har medfört vissa svårigheter. Det finns till exempel argument både för och emot att metoden genererar data som är lokalt oberoende, det vill säga att varje bedömning är opåverkad av vilka andra bedömningar som har gjorts (Debelak & Koller, 2020) och därmed lämplig för statistisk analys (se till exempel Baghaei, 2007; Bramley, 2007; Marais & Andrich, 2008). Författarna har utgått från den forskning som använt och utvärderat metoden (se till exempel Boonen m.fl., 2020; Stone m.fl., 2020) men noterar en problematik i att det saknas konsensus kring hur datan ska tolkas och användas.

Även de lingvistiska mått som användes i föreliggande studie har begränsningar. Bland annat är det till författarnas kännedom okänt hur väl måtten, undantaget MTLD, fungerar för olika textlängder. I föreliggande studie fick till exempel ett fåtal texter mycket hög syntaktisk komplexitet på grund av att de bestod av ett fåtal t-units med många bisatser. Detta bör emellertid inte ha påverkat resultatet nämnvärt i och med att den icke-parametriska statistiken som användes borde kompensera för detta. Utöver detta påverkas korrekthetsmåtten till viss del av den syntaktiska komplexiteten. Detta eftersom två texter med lika många ord och fel kan få olika andel korrekta t-units beroende på hur många och hur långa t-units de innehåller. En lösning på detta hade varit att istället räkna antal fel per ord. Detta valdes dock bort eftersom det i många fall var svårt att avgöra var ett fel slutade och ett annat började. Det fanns också en takeffekt för korrekthetsmåtten, eftersom många texter saknade fel. Detta kan ha gjort det svårare att få signifikanta resultat på korrelationsanalyserna. Utöver detta är det viktigt att notera att den bedömda textkvaliteten kan ha påverkats av lingvistiska variabler som inte undersöktes i föreliggande studie. Exempel på detta är kohesion, texternas innehåll samt hur väl deltagarna följt uppgiften.

Avslutningsvis är texterna som bedömts och analyserats i föreliggande studie producerade under omständigheter som till viss del skiljer sig från de som råder i barnens vardagliga skrivsituationer. Utöver att barnen hade marginell erfarenhet av att skriva med taligenkänning, fick de till exempel inte använda stavningskontroll. Därutöver korrigerades stav- och taligenkänningsfel inför bedömningen vilket inte skulle ha skett i en klassrumssituation. Hur dessa förutsättningar påverkade texterna är oklart men behöver beaktas när man tolkar resultaten.

## *Kliniska implikationer*

Utifrån resultaten från föreliggande studie kan det inte antas att taligenkänningen fungerar som hjälpmedel för alla barn med stavningssvårigheter. Om interventionen ska vara framgångsrik krävs förmodligen både träning i hur man använder taligenkänningsverktyg

och att detta kombineras med annan skrivträning anpassad efter individens styrkor och svårigheter. Eftersom det i föreliggande studie fanns stor variation i utfall är det också viktigt att man för varje barn utvärderar effekten av hjälpmedlet.

### *Framtida forskning*

Ytterligare studier krävs för att klargöra hur skrivande med taligenkänning skiljer sig från traditionella skrivsätt. Det finns ett fortsatt behov av att undersöka huruvida lingvistiska egenskaper skiljer texter skrivna med taligenkänning från traditionellt skrivna texter. Även om inga sådana skillnader fanns i föreliggande studie, är det möjligt att det finns egenskaper, exempelvis kohesion, som är svårare att uppnå i något av skrivsätten. En kvalitativ textanalys skulle kunna ge fördjupad kunskap om vad som utmärker skrivsätten samt vad som krävs för att en text ska bedömas ha hög textkvalitet. Experimentella studier skulle kunna ge information om hur den relativa belastningen av olika processer, till exempel transkription och exekutiva funktioner, skiljer sig mellan skrivsätten. Detta skulle klargöra ifall kraven på exekutiva funktioner är högre när man skriver med taligenkänning. Ur ett kliniskt perspektiv är det också viktigt att undersöka huruvida särskilda språkliga eller kognitiva förmågor, såsom avkodning, språkförståelse, uppmärksamhet och arbetsminne, påverkar möjligheten att dra nytta av taligenkänning. Detta eftersom olika skrivhjälpmedel och interventioner sannolikt fungerar olika väl för olika individer. För en del kan taligenkänning tänkas vara det bästa alternativet, för andra kanske stavningskontroll eller ordprediktion fungerar bättre och för ytterligare andra kan en kombination vara det som ger mest stöd i skrivandet. Avslutningsvis är det angeläget att undersöka hur skrivande med taligenkänning utvecklas över tid och med erfarenhet. Interventionsstudier, där barn ges explicit undervisning och träning i att skriva med taligenkänning, skulle därför kunna ge en mer omfattande bild av verktygets potential för barn med stavningssvårigheter.

### Referenser

- Anthony, L. (2012). *AntWordProfiler* (Version 1.5.0, Windows) [Programvara]. <https://www.laurenceanthony.net/software/antwordprofiler/>
- Arcon, N., Klein, P. D., & Dombroski, J. D. (2017). Effects of dictation, speech to text, and handwriting on the written composition of elementary school English language learners. *Reading and Writing Quarterly*, 33(6), 533-548. <https://doi.org/10.1080/10573569.2016.1253513>
- Baghaei, P. (2007). Local dependency and Rasch measures. *Rasch Measurement Transactions*, 21, 1105-1106.
- Berninger, V. W., Vaughan, K., Abbott, R. D., Begay, K., Coleman, K. B., Curtin, G., Hawkins, J. M., & Graham, S. (2002). Teaching spelling and composition alone and together: Implications for the Simple View of Writing. *Journal of Educational Psychology*, 94(2), 291-304. <https://doi.org/10.1037/0022-0663.94.2.291>
- Bishop, D. V. M., & Clarkson, B. (2003). Written language as a window in to residual language deficits: A study of children with persistent and residual speech and

- language impairments. *Cortex*, 39(2), 215-237. [https://doi.org/10.1016/S0010-9452\(08\)70106-0](https://doi.org/10.1016/S0010-9452(08)70106-0)
- Boonen, N., Kloots, H., & Gillis, S. (2020). Rating the overall speech quality of hearing-impaired children by means of comparative judgements. *Journal of communication disorders*, 83. <https://doi.org/10.1016/j.jcomdis.2019.105969>
- Bourdin, B., & Fayol, M. (1994). Is written language production more difficult than oral language production? A working memory approach. *International Journal of Psychology*, 29(5), 591-620. <https://doi.org/10.1080/00207599408248175>
- Bramley, T. (2007). Paired comparison methods. I P. Newton, J. Baird, H. Goldstein, H. Patrick, & P. Tymms (Red.), *Techniques for monitoring the comparability of examination standards* (s. 246-294). [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/487059/2007-comparability-exam-standards-i-chapter7.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/487059/2007-comparability-exam-standards-i-chapter7.pdf)
- Brimo, K., Dinkler, L., Gillberg, C., Lichtenstein, P., Lundström, S., & Åsberg Johnels, J. (2021). The co-occurrence of neurodevelopmental problems in dyslexia. *Dyslexia*, 1-17. <https://doi.org/10.1002/dys.1681>
- Caravolas, M., Hulme, C., & Snowling, M. J. (2001). The foundations of spelling ability: Evidence from a 3-year longitudinal study. *Journal of memory and language*, 45(4), 751-774. <https://doi.org/10.1006/jmla.2000.2785>
- Crossley, S. (2020). Linguistic features in writing quality and development: An overview. *The Journal of Writing Research*, 11, 415-443.
- Debelak, R., & Koller, I. (2020). Testing the local independence assumption of the Rasch model with Q3-based nonparametric model tests. *Applied Psychological Measurement*, 44(2), 103-117. <https://doi.org/10.1177/0146621619835501>
- Dockrell, J. (2009). Causes of delays and difficulties in the production of written text. I D. M. R. Beard, J. Riley & M. Nystrand (Red.), *The SAGE handbook of writing development* (s. 489-505). SAGE.
- Graham, S., Harris, K., & Hebert, M. (2011). It is more than just the message: Presentation effects in scoring writing. *Focus on Exceptional Children*, 44(4), 1-12. <https://doi.org/10.17161/fec.v44i4.6687>
- Graham, S., Harris, K. R., & Adkins, M. (2018). The impact of supplemental handwriting and spelling instruction with first grade students who do not acquire transcription skills as rapidly as peers: a randomized control trial. *Reading and Writing*, 31(6), 1273-1294. <https://doi.org/10.1007/s11145-018-9822-0>
- Hatcher, J., Snowling, M. J., & Griffiths, Y. M. (2002). Cognitive assessment of dyslexic students in higher education. *British Journal of Educational Psychology*, 72, 119-133. <https://doi.org/https://doi.org/10.1348/000709902158801>
- Hayes, J., & Flower, L. (1980). Identifying the organization of writing processes. I G. L. W. & Steinberg E. R. (Red.), *Cognitive processes in writing: An interdisciplinary approach* (s. 3-30). Erlbaum. [https://www.researchgate.net/publication/200772468\\_Identifying\\_the\\_organization\\_of\\_writing\\_processes](https://www.researchgate.net/publication/200772468_Identifying_the_organization_of_writing_processes)
- Hunt, K. W. (1965). *Grammatical structures written at three grade levels*. NCTE research report no. 3. <https://files.eric.ed.gov/fulltext/ED113735.pdf>
- Johansson, V. (2009). *Developmental aspects of text production in writing and speech* [Doktorsavhandling]. Lunds universitet. <https://lup.lub.lu.se/search/publication/1487257>

- Järpsten, B., & Taube, K. (2010). *DLS för skolår 4-6.Handledning*. Hogrefe Psykologiförlaget.
- Katusic, S. K., Colligan, R. C., Weaver, A. L., & Barbaresi, W. J. (2009). The forgotten learning disability: epidemiology of written-language disorder in a population-based birth cohort (1976-1982), Rochester, Minnesota. *Pediatrics*, *123*(5), 1306-1313. <https://doi.org/10.1542/peds.2008-2098>
- Kraft, S., Thurfjell, F., Rack, J., & Wengelin, Å. (2019). Lexikala analyser av muntlig, tangentbordsskriven och dikterad text producerad av barn med stavningssvårigheter. *Nordic Journal Of Literacy Research*, *5*(3). <https://doi.org/10.23865/njlr.v5.1511>
- Landerl, K., & Moll, K. (2010). Comorbidity of learning disorders: prevalence and familial transmission. *Journal of Child Psychology and Psychiatry*, *51*(3), 287-294. <https://doi.org/10.1111/j.1469-7610.2009.02164.x>
- Leijten, M. (2007). *Writing and speech recognition : Observing error correction strategies of professional writers* [Doktorsavhandling]. LOT-school Utrecht.
- Lix. (2021). [Onlinebaserat verktyg för lingvistisk analys]. Hämtad 18 feb från <http://www.lix.se/>
- MacArthur, C. A. (2009). Reflections on research on writing and technology for struggling writers. *Learning Disabilities Research & Practice*, *24*(2), 93-103. <https://doi.org/10.1111/j.1540-5826.2009.00283.x>
- MacArthur, C. A., & Cavalier, A. R. (2004). Dictation and speech recognition technology as test accommodations. *Exceptional children*, *71*(1), 43-58. <https://doi.org/10.1177/001440290407100103>
- Marais, I., & Andrich, D. (2008). Formalizing dimension and response violations of local independence in the unidimensional Rasch model. *Journal of Applied Measurement*, *9*(3), 200-215.
- Marshall, J. C., & Powers, J. M. (1969). Writing neatness, composition errors, and essay grades. *Journal of Educational Measurement*, *6*(2), 97-101. <https://doi.org/https://doi.org/10.1111/j.1745-3984.1969.tb00665.x>
- Maughan, B., Messer, J., Collishaw, S., Pickles, A., Snowling, M., Yule, W., & Rutter, M. (2009). Persistence of literacy problems: spelling in adolescence and at mid-life. *Journal of Child Psychology and Psychiatry*, *50*(8), 893-901. <https://doi.org/10.1111/j.1469-7610.2009.02079.x>
- McCarthy, P. M., & Jarvis, S. (2010). MTL, D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*(2), 381-392. <https://doi.org/10.3758/BRM.42.2.381>
- McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychology Review*, *8*(3), 299-325. <https://doi.org/10.1007/BF01464076>
- McHale, K., & Cermak, S. A. (1992). Fine motor activities in elementary school: preliminary findings and provisional implications for children with fine motor problems. *The American Journal of Occupational Therapy*, *46*(10), 898-903. <https://doi.org/10.5014/ajot.46.10.898>
- Moll, K., Kunze, S., Neuhoff, N., Bruder, J., & Schulte-Körne, G. (2014). Specific learning disorder: Prevalence and gender differences. *PLoS ONE*, *9*(7), 1-8. <https://doi.org/10.1371/journal.pone.0103537>



- Moll, K., & Landerl, K. (2009). Double dissociation between reading and spelling deficits. *Scientific Studies of Reading*, 13(5), 359-382.  
<https://doi.org/10.1080/10888430903162878>
- Mukaka, M. M. (2012). Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi medical journal : the journal of Medical Association of Malawi*, 24(3), 69-71.
- Nelson, L. M., & Reynolds, T. W., Jr. (2015). Speech recognition, disability, and college composition. *Journal of Postsecondary Education and Disability*, 28(2), 181-197.
- No More Marking. (2021). [Onlinebaserat verktyg för comparative judgement]. Hämtad 2021-02-05 från [www.nomoremarking.com](http://www.nomoremarking.com)
- Nyström, C. (2000). *Gymnasisters skrivande : en studie av genre, textstruktur och sammanhang* [Doktorsavhandling]. Uppsala universitet.
- O'Shaughnessy, D. (2003). Interacting with computers by voice: automatic speech recognition and synthesis. *Proceedings of the IEEE*, 91(9), 1272-1305.  
<https://doi.org/10.1109/JPROC.2003.817117>
- Olive, T., & Kellogg, R. T. (2002). Concurrent activation of high- and low-level production processes in written composition. *Memory & Cognition*, 30(4), 594-600. <https://doi.org/10.3758/BF03194960>
- Peterson-Karlan, G. R. (2011). Technology to support writing by students with learning and academic disabilities: Recent research trends and findings. *Assistive Technology Outcomes and Benefits*, 7, 39-62.  
<https://files.eric.ed.gov/fulltext/EJ961161.pdf>
- Pollitt, A. (2012). Comparative judgement for assessment. *International Journal of Technology and Design Education*, 22(2), 157-170.  
<https://doi.org/10.1007/s10798-011-9189-x>
- Quinlan, T. (2004). Speech recognition technology and students with writing difficulties: Improving fluency. *Journal of Educational Psychology*, 96(2), 337-346. <https://doi.org/10.1037/0022-0663.96.2.337>
- Santangelo, T., & Graham, S. (2016). A comprehensive meta-analysis of handwriting instruction. *Educational Psychology Review*, 28(2), 225-265.  
<https://doi.org/10.1007/s10648-015-9335-1>
- Scardamalia, M., & Bereiter, C. (1987). *The psychology of written composition*. Erlbaum.  
[https://books.google.se/books?hl=en&lr=&id=zdP5AQAAQBAJ&oi=fnd&pg=PI&ots=d\\_NWtMxEYT&sig=zPybGyldlQ3TX2577nQas-YZENA&redir\\_esc=y#v=onepage&q&f=false](https://books.google.se/books?hl=en&lr=&id=zdP5AQAAQBAJ&oi=fnd&pg=PI&ots=d_NWtMxEYT&sig=zPybGyldlQ3TX2577nQas-YZENA&redir_esc=y#v=onepage&q&f=false)
- Skolinspektionen. (2011). *Lika eller olika? Omrättning av nationella prov i grundskolan och gymnasieskolan*.  
<https://www.skolinspektionen.se/globalassets/02-beslut-rapporter-stat/granskningsrapporter/tkg/2011/lik-eller-olika-omratt2011-slutrapport.pdf>
- Skolverket. (2009). *Bedömaröverensstämmelse vid bedömning av nationella prov*.  
<https://www.skolverket.se/download/18.6bfaca41169863e6a6585e0/1553961864204/Bed%C3%B6mar%C3%B6verensst%C3%A4mmelse%20vid%20bed%C3%B6mning%20av%20nationella%20prov.pdf>
- Skolverket. (2019). *PM - Resultat på nationella prov i årskurs 3, 6 och 9, läsåret 2018/19*. [https://siris.skolverket.se/siris/sitevision\\_doc.getFile?p\\_id=549421](https://siris.skolverket.se/siris/sitevision_doc.getFile?p_id=549421)

- Statens offentliga utredningar. (2017). *Samling för skolan. Nationell strategi för kunskap och likvärdighet*. [http://www.sou.gov.se/wp-content/uploads/2017/04/SOU\\_2017\\_35\\_webb\\_Final-version.pdf](http://www.sou.gov.se/wp-content/uploads/2017/04/SOU_2017_35_webb_Final-version.pdf)
- Steedle, J. T., & Ferrara, S. (2016). Evaluating Comparative Judgment as an approach to essay scoring. *Applied Measurement in Education*, 29(3), 211-223. <https://doi.org/10.1080/08957347.2016.1171769>
- Stone, G., Andrade, J., Martin, K., & Styles, B. (2020). *Helping handwriting shine*. The Education Endowment Foundation. [https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation\\_Reports/Helping\\_Handwriting\\_Shine\\_Report\\_-\\_final.pdf](https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Helping_Handwriting_Shine_Report_-_final.pdf)
- Swegram. (2021). [Webbaserat verktyg för annotering och analys av svenska texter]. Hämtad 2021-03-03 från <https://cl.lingfil.uu.se/swegram/>
- Teleman, U. (1999). *Svenska akademiens grammatik* (1 uppl.). Norstedts.
- Textinspector. (2021). [Onlinebaserat verktyg för lexikal analys]. Hämtad 2021-02-18 från [www.textinspector.com](http://www.textinspector.com)
- van Daal, T., Lesterhuis, M., Coertjens, L., Donche, V., & De Maeyer, S. (2016). Validity of comparative judgement to assess academic writing: Examining implications of its holistic character and building on a shared consensus. *Assessment in Education Principles Policy and Practice*, 26, 59-74. <https://doi.org/10.1080/0969594X.2016.1253542>
- Van Steendam, E., Tillema, M., & Rijlaarsdam, G. (2012). Introduction. I E. V. Steendam (Red.), *Measuring writing: Recent insights into theory, methodology and practice* (s. ix-xxi). BRILL. <https://doi.org/10.1163/9789004248489>
- Verhavert, S., Bouwer, R., Donche, V., & De Maeyer, S. (2019). A meta-analysis on the reliability of comparative judgement. *Assessment in Education: Principles, Policy & Practice*, 26(5), 541-562. <https://doi.org/10.1080/0969594X.2019.1602027>
- Verhavert, S., De Maeyer, S., Donche, V., & Coertjens, L. (2018). Scale separation reliability: What does it mean in the context of Comparative Judgment? *Applied Psychological Measurement*, 42(6), 428-445. <https://doi.org/10.1177/0146621617748321>
- Weadon, C. (2018, 19 feb). Judge infit. *The No More Marking Blog*. <https://blog.nomoremarking.com/judge-infit-27aec5ede2d>
- Wells Rowe, D. (2009). Early written communication. I R. Beard, D. Myhill, M. Nystrand, & J. Riley (Red.), *The SAGE handbook of writing development* (s. 213-231). SAGE Publications Ltd.

## Bilaga 1

Vid det första tillfället testades deltagarnas språkförståelse (TROG-2), arbetsminne (CLPT del 1), avkodningsförmåga (LäSt), RAN (Logos), och läsförståelse (Logos). De fick även svara på frågor gällande läs- och skrivvanor. Därefter fick deltagarna se en av två filmer och hålla ett muntligt föredrag utifrån uppgiften:

*”Du har nu sett en filmsnutt där någon person ställs inför något slags problem. Jag är intresserad av att veta vad folk tänker om detta! Tänk om det skulle komma en superhjärte och se vad som händer! Vad skulle den tycka då? Du får välja en superhjärte som redan finns eller hitta på en egen. Vad tror du att din superhjärte skulle ha att säga om det här? Vi samlar in uppsatser och föredrag/föreläsningar/presentationer om detta och jag vill att du förklarar vad du tror att din superhjärte skulle tycka om situationen. Beskriv inte bara filmen, utan beskriv också hur din superhjärte skulle resonera och diskutera.”*

Efter detta fick de återigen se en av de två filmerna och svara på samma uppgift som tidigare. Denna gång skulle de skriva ner svaret, antingen med tangentbord eller med taligenkänning. Vid det andra tillfället testades deltagarnas hörförståelse (Logos), arbetsminne (CLTP del 2), ordförråd (CELF likheter 2), stavningsförmåga (DLS) och ordmobiliseringsförmåga (CELF). Därefter fick deltagarna se en av filmerna och skriftligen svara på uppgiften med tangentbord eller taligenkänning. Efter detta fick de återigen se en av filmerna och svara på uppgiften, med tangentbord eller med taligenkänning. Vilket filmklipp deltagarna fick se samt vilket skrivverktyg som skulle användas för respektive textproduktion slumpades, för att minska risken att ordningsföljden skulle påverka resultatet. Deltagarna testades och skrev texterna enskilt med en av logopederna. För information om ytterligare instruktioner, introduktion av dikteringsverktyget och inspelningsförfarande hänvisas till Kraft m.fl. (2019).