

K-12 教育背景下文本難度研究的系統性綜述（1948-2024）

張興利 李宜遜* 邱智敏

摘要

本研究着眼於學生閱讀能力發展及其促進，基於閱讀學習的視角，聚焦 K-12 教育背景下閱讀材料（教材、教輔及課外讀物）的文本難度，旨在通過理論指導實踐，助力教育系統在基礎教育階段培養能夠理解複雜文本的讀者，為其後續的高等教育和社會生活奠定紮實的閱讀基礎。因此，本研究採用系統性文獻綜述法，從研究特徵、評估方法、影響因素、支持策略四方面，對 187 篇有關 K-12 階段文本難度的實證研究文獻進行系統梳理和內容分析，為我國相關領域的研究提供參考。結果發現：（1）以往研究主要聚焦於文本語言複雜性、英語教材難度及基於大型語料庫來分析文本難度；研究學段則集中在小學和中學階段，少有對幼稚園階段文本難度的分析；研究材料以英語類教材為主，對非英語文本的探索有限。（2）採用可讀性公式的定量評估是評定文本難度的主要方法，尤其是在小學和中學階段。（3）不同類型的文本特徵是文本難度的主要影響因素。（4）較少研究關注促進學生閱讀複雜文本的策略。

關鍵詞：K-12 教育 文本難度 系統性文獻綜述 實證研究 內容分析

張興利，香港教育大學幼兒教育學系，聯絡電郵：s1149074@s.edu.hk

* 李宜遜，香港教育大學幼兒教育學系，聯絡電郵：lyixun@eduhk.hk（本文通訊作者）

邱智敏，香港教育大學幼兒教育學系，聯絡電郵：s1140789@s.edu.hk

本文系香港教育大學研究與發展副校長辦公室“兒童雙語閱讀發展”課題（#04753）階段性研究成果

一、研究背景

閱讀能力是學生學業發展的基石。主流的閱讀發展階段論認為，個體的閱讀技能在系統教學中逐步、持續發展，呈現出一定的發展階段（Chall, 1983）。從幼稚園階段的接觸閱讀，到小學低年級「學習閱讀」（Learning to Read），至四年級後「在閱讀中學習」（Reading to Learn），即學生通過閱讀文本來學習不同科目，亦在大量閱讀中進一步提升閱讀水準（Harlaar, 2007）。隨着年級的增加，學生閱讀文本的內容也更為複雜，這要求學生具備更高的閱讀能力，既能支持其從複雜文本中汲取新知識，又可避免因文本難度過高而無法進行資訊加工。

文本難度（Text Difficulty）指書面文本被閱讀加工的難易程度，亦稱文本複雜度（Text Complexity）或文本易讀度（Text Readability）（Mesmer et al., 2012）。文本難度受多重因素的影響，主要包括文本本身的特徵（如單詞長度、句法複雜性）、閱讀者的特點（如背景知識、閱讀動機）和閱讀任務的特點（如任務要求、問題類型）。當文本包含更多的複雜文本特徵（如複雜句），則其難度更高；同一文本對不同讀者而言閱讀難度亦有所區別，讀者的個體差異會影響其理解文本的難度，如具備更多背景知識的讀者更容易理解文本；此外，同一文本對同一讀者而言，其難度仍可能因閱讀任務的變化而不同，如教師講解文本可降低閱讀難度，而獨立閱讀則要求個體更深層次加工，這增加了閱讀難度。因此，閱讀研究領域廣泛認為，文本、讀者和任務均為閱讀理解的重要組成部分（Snow, 2002）。在研究文本難度時，應綜合考慮多方因素。因此，本研究聚焦 K-12（幼稚園至高中）教育背景下的文本難度研究，既囊括文本、讀者和任務三方因素，亦探討不同年級文本難度的評估方法及支持讀者閱讀複雜文本的有效策略。本研究旨在基於閱讀學習的視角（Alexander & Jetton, 2000），理論指導實踐，助力教育系統在基礎教育階段（K-12）培養能夠理解複雜文本的讀者，為其後續接受高等教育或參與社會工作生活奠定紮實的閱讀能力基礎（Fitzgerald et al., 2015）。

國內關於 K-12 階段文本難度研究主要聚焦於漢語語言體系，尤

其關注語文教材文本難度的評估。實際上，關於漢語文本難度的研究最早可追溯到美國威斯康辛麥迪遜大學博士論文 Yang (1971)，其以漢語文本為材料，分析字、詞、句等文本特徵，以高中生閱讀理解成績標定文本難度，探索了文本特徵對漢語文本難度的影響，並通過線性回歸分析建立了漢語可讀性公式。後續研究也多以探討漢語文本難度的影響因素，並嘗試建立文本難度評估標準（可讀性公式）為主。系統檢索顯示¹，該領域至少有十篇研究成果。例如，邊偉等（2017）通過學生的主觀評定來標定小學高年級語文教科書文本的難度，發現其受學生性別、文本主題和類型的影響。張秋玲等（2022）則人工編碼了中學語文教科書文言文的詞彙和語序特徵，建立了文言文難度等級（由學生閱讀能力及專家評定所標定）的指標參照值。此外，七篇研究在鑒別對文本難度有重要影響的文本特徵的基礎之上，亦建立了這些文本特徵與文本難度之間的統計預測模型，以此來構建可讀性公式；語料庫均為語文教材（小學或中小學），文本難度的標定均參照文本出現在教科書的年級段（學段、年級或學期）。其中兩篇研究採用經典統計方法（程勇等，2020；劉苗苗等，2021），另五篇研究則基於自然語言處理 (Natural Language Processing, NLP) 和機器學習演算法 (Machine Learning Algorithms)，實現文本難度的自動化評估（程勇、徐德寬、董軍，2020；程勇、徐德寬、呂學強，2019；李文彪、吳雲芳，2023；吳思遠等，2020；柏曉鵬、吉伶俐，2022）。相應地，國內已有三篇關於文本難度研究的綜述均聚焦於文本的可讀性評估。王蕾（2022）分析了可讀性公式的發展階段及特點，並討論了不同階段可讀性公式的研究範式和代表性研究成果等。吳思遠等（2018）介紹了文本可讀性自動分析的方法、文本特徵的選擇及相關資料庫。劉瀟（2015）則梳理了國內外有關英語、法語和漢語可讀性公式的發展階段和特點。整體而言，目前國內關於 K-12 階段文本難度的研究主要聚

1 以「文本難度」、「文本複雜性」、「可讀性公式」、「文本易讀度」為關鍵字在中國知網資料庫進行系統性文獻檢索。不限制開始時間，截止時間為 2024 年 4 月。檢索語言限定為中文，文獻類型為核心期刊（北大核心、CSSCI），納入標準為實證研究、全文可獲取、K-12 階段教育類文本（教材、教輔資料及課外讀物）。

焦漢語文本，側重於文本特徵及其對文本難度的影響，較少考慮讀者和任務因素，且缺乏對閱讀支持策略的關注。

相較於國內對漢語文本難度的集中探討，國外的文本難度研究起步更早，範圍廣泛且深入，覆蓋了更豐富的語言體系。其研究內容不僅涉及文本特徵，亦考慮讀者和任務因素對文本難度的影響。此外，還探討了針對複雜文本的閱讀支持策略。首先，部分研究與國內相似，考慮單一的文本特徵因素對文本難度的影響。例如，Ghani et al. (2014) 將中學教材作為樣本，分析文本特徵（單詞、句子等）對文本難度的預測效果，採用經典統計方法構建預測模型，開發阿拉伯語文本的可讀性公式；Balyan et al. (2020) 則基於文本特徵因素，用自然語言處理和機器學習演算法來估計英語教材的文本難度。其次，一些研究則同時考慮兩個以上因素對文本難度的影響。例如，Mikk & Elts (1999) 以高中科學教材為語料庫，綜合考慮文本特徵（詞長、句長及句法）和讀者特點（背景知識、閱讀興趣及文本熟悉度）來建立俄語文本的可讀性公式；Spencer et al. (2019) 依據項目反應理論（Item Response Theory），評估中小學生閱讀理解測試結果，並分析讀者特點（詞彙知識、認知技能）和任務因素（不同問題類型）對讀者理解複雜文本的影響。再者，有研究探討文本難度的支持策略。例如，Fisher & Frey (2014) 提出細讀（Close Reading）的閱讀策略，以支持學生閱讀複雜文本。可以預見，隨着國內學界對文本難度研究的進一步關注，未來研究很可能會拓展到結合文本、讀者和任務多方因素，深化對漢語文本難度的理解。因此，有必要對國外研究進行系統回顧和總結，期望為非英語類文本尤其是漢語文本難度的研究提供參考。

二、研究設計與實施

（一）研究問題

本研究依據系統性綜述法 (Systematic Review)，涵蓋截止至 2024 年 4 月、K-12 教育背景下關於文本難度的實證研究，圍繞研究問題，

結合文獻計量分析和內容分析法，對現有研究特徵、文本難度的評估方法、影響因素、支持策略等方面進行分析。本研究共有四個待答問題：

1. 以往研究呈現哪些特點（研究趨勢、研究對象）？
2. 有哪些方法可以評估 K-12 階段的文本難度？
3. 有哪些因素影響 K-12 階段的文本難度？
4. 有哪些教學策略可以有效支持學生閱讀複雜文本？

（二）研究方法

本研究採用系統性文獻綜述法對 K-12 教育中文本難度的相關實證研究進行審查。圍繞上述研究問題，事先制定文獻納入和排除標準（見表 1），並使用系統和標準化的程式來識別、選擇和評價相關研究（Siddaway et al., 2019）。同時，為深入挖掘 K-12 教育文本難度的核心議題，本研究將對最終納入審查的文獻進行內容分析（White & Marsh, 2006）。此外，為了解 K-12 教育文本難度研究領域的發展概況，本研究使用 CiteSpace 軟件對文獻進行計量分析（Chen, 2006）。

（三）樣本選取

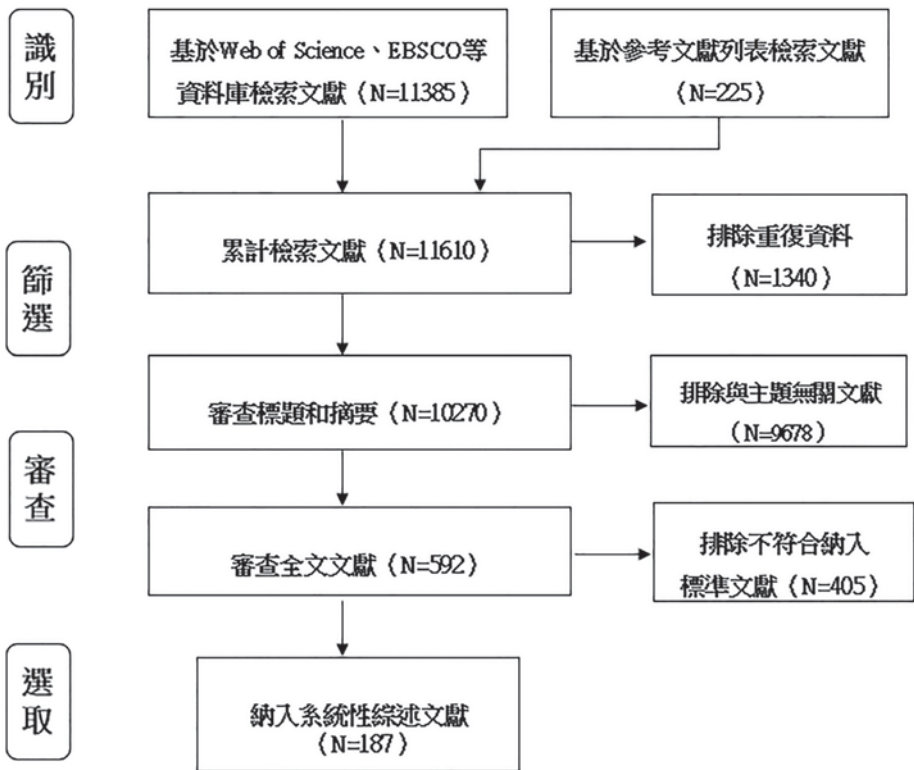
本研究採用粗略關鍵詞檢索，以“Text Complexity”、“Text Difficulty”、“Text Readability”、“Text Characteristics”、“Text Feature”、“Text Challenge”、“Reading Complexity”、“Reading Comprehension”英文關鍵詞在常用的六個英文教育學資料庫 EBSCO、Web of Science、ProQuest、Scopus、SAGE Journal、Spring Link 中進行檢索，並對檢索文獻的參考文獻列表進行篩查。其中“Text”統一指代全學段（K-12）所涉及的多樣化閱讀文本材料，包括圖畫書（Picture Book）、教科書、教輔資料、課外讀物及短篇文章等。檢索時間為 2024 年 4 月，檢索語言為英文，文獻類型為同行評審期刊論文，依據檢索到的全文文獻中最早的發表時間，確定時間範圍為 1948-2024 年 4 月，初步檢索 11,610 篇文獻。

基於文獻納入和排除標準（見表 1），結合 PRISMA（Preferred Reporting Items for Systematic Reviews and Meta-Analyses）的文獻篩選框架（見圖 1）（Page et al., 2021），本研究對檢索文獻的標題及摘要進行第一輪篩選（N=592），再對文獻全文進行第二輪篩選，最終得到 187 篇文獻納入分析。

表 1：文獻納入與排除標準

序號	納入標準	排除標準
1	實證研究	非實證研究，如綜述或觀點評論類文獻
2	同行評審期刊論文	非同行評審期刊論文，如會議、學位論文等
3	全文可獲取	非全文獲取
4	英文論文	非英文論文
5	K-12 教育學段	非 K-12 教育學段，如大學高等教育和成人職業教育
6	教育類文本（教材、教輔及課外讀物）	非教材、教輔及課外讀物，如學生手冊、家長同意書或藥物說明書等

圖 1：PRISMA 文獻篩選流程圖



(四) 文獻編碼

為進一步對納入文獻進行內容分析，本研究建立了系統的文獻編碼體系（見表 2）。同時，為確保編碼的可靠性，第一作者與第三作者對隨機抽取的 30% 文獻分別進行獨立編碼，評分者一致性為 85.4%，表明編碼程式高度可信。兩位作者亦經過討論，對不一致的編碼達成一致意見，並形成最終的編碼分析框架。在此基礎上，第一作者對餘下樣本進行全部編碼。

表 2：K-12 教育文本難度實證研究內容分析編碼框架

一級維度	二級維度	三級維度
研究特徵	研究趨勢	發文數量、研究熱點
	研究對象	研究學段、研究學科、文本類型、目標語言
評估方法	方法類別	可讀性公式評估、專家評估、讀者與任務評估
影響因素	因素類別	文本、讀者、任務
支持策略	策略類型	教學策略、閱讀策略

三、研究結果與分析

(一) 現有研究特徵

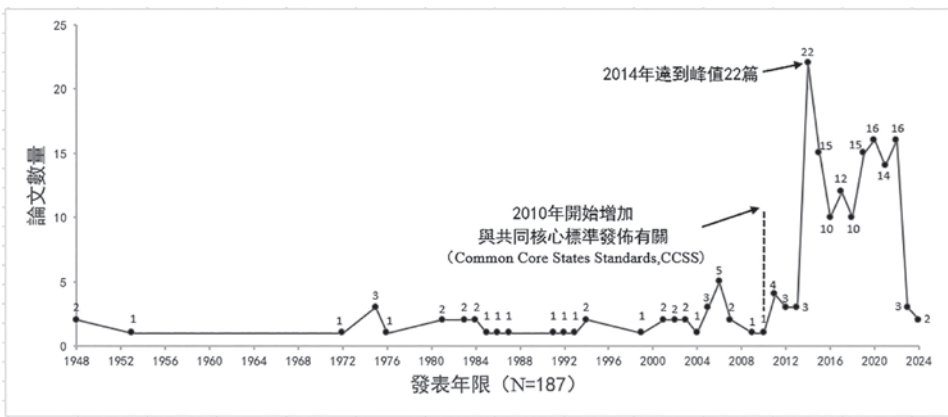
1. 研究趨勢

從發文數量看，過去 76 年間（1948-2024），文本難度研究發文數量總體呈現波狀趨勢（見圖 2）。發文數量自 2010 年起明顯增加，並於 2014 年達到峰值。這可能與 2010 年美國頒佈的共同核心國家標準（Common Core State Standards, CCSS）² 有關（Bomer & Maloch,

2 美國《共同核心國家標準》（Common Core State Standards, CCSS），由全美州長協會（NGA）和州立學校官員委員會（CCSSO）於 2010 年聯合制定，最初 48 個州共同參與，旨在解決各州教育標準差異過大的問題。截至 2024 年，41 個州及哥倫比亞特區仍在採用。CCSS 主要涵蓋英語語言藝術（ELA）和數學領域。在英語語言藝術方面，着重對 K-12 學生分析複雜文本及閱讀能力提出要求，並構建了包含定量指標、定性分析及讀者任務情境的文本難度評估框架。

2011)。該標準釐定了學生在 K-12 階段需達到的閱讀能力要求，並介紹了閱讀材料文本難度的評估方法，引發了學界與教育界對文本難度的廣泛討論（Gamson et al., 2013）。如美國芝加哥大學出版的核心學術期刊《小學學報》（*The Elementary School Journal*）於 2014 年組織特刊聚焦文本難度研究議題。儘管 2014 年後發文量有所下降，但仍保持在每年 13 篇左右，表明研究者對此話題的持續關注。2024 年以後發文量的下降可能與當年發表的文章尚未被納入數據庫有關，而非研究熱度的下降。

圖 2：1948-2024 年 K-12 教育文本難度實證研究發文數量



從研究趨勢看，本研究借助 CiteSpace 中的 Burst Term 功能對近十年來（2014-2024）的關鍵詞進行突現分析（Keyword Burst Analysis），以了解 K-12 教育文本難度實證研究領域的最新研究熱點（見表 3）。首先，過去十年間僅出現 8 個突現詞，少量突現詞表明學界對該領域的研究視角相對集中。其次，「語言複雜性」（強度為 1.8）和「英語教科書」（強度為 1.6）為突現強度相對較大的關鍵詞，表明已有研究更多聚焦在這兩方面。再者，自 2019 年以來，突現的關鍵詞除了「語言複雜性」和「英語教科書」，還有「文本難度」（強度為 1.15）、「自然語言處理」（強度為 0.94）和「語料庫語言學」（強度為 0.94），體現了近幾年該領域研究對新技術和演算法的應用。隨着計算

機技術和人工智能的進步，大規模語料庫的構建變得更加可行，機器學習相關演算法和數據統計軟件也在不斷優化，這些都為文本難度的研究提供了便利。因此，關注文本語言複雜性、聚焦英語教科書文本難度、借助自然語言處理技術評估文本難度、利用大型語料庫分析文本難度是未來 K-12 教育文本難度領域的重要趨勢。

表 3：不同時期 K-12 教育文本難度實證研究的八個關鍵詞突現情況（2014-2024）

突現詞	突現強度	開始年份	截止年份	2014-2024
複雜文本（Complex Texts）	1.05	2015	2016	
小學生（Elementary Student）	0.84	2016	2018	
困難文本（Difficult Texts）	0.84	2016	2018	
語言複雜性（Linguistic Complexity）	1.8	2018	2020	
文本難度（Text Difficulty）	1.15	2019	2021	
自然語言處理（Natural Language Processing）	0.94	2020	2021	
英語教科書（English Textbooks）	1.6	2021	2024	
語料庫語言學（Corpus Linguistics）	0.94	2021	2024	

注：突現詞指特定時間段出現頻率顯著增加的關鍵詞；突現強度（數值）越高，表明該關鍵詞在文獻中出現頻率越高。

2. 研究對象

從研究學段看（見表 4），目前關於文本難度的研究呈現明顯的學段分佈不均：主要集中在小學和中學階段，佔比達 91.4%（N=171），而幼稚園階段僅佔 8.6%（N=16）。值得注意的是，本研究已完整納入幼稚園學段的圖畫書研究，但數據仍顯示該領域文獻較少。這可能與不同年齡層的文本特點及其與現有的文本難度評估方法和技術的適配度有關。小學和中學的文本通常包含較多的文字資訊，更適合使用文本難度領域中發展較為成熟的量化評估技術，例如各類可讀性公式或內嵌可讀性公式的平台或軟件可對文本的難度進行自動化分析，研究的便利性較高。而幼稚園階段的文本通常文字較少，更多涉及圖畫、顏色、排版等多模態特徵，能夠進行自動化量化分析的因素較少，通

常會涉及對文本難度的定性評估，對評估框架的成熟度和參與人工評定的專家等都有更高的要求，這可能在一定程度上限制了研究的可行性。

表 4：研究文獻的學段分佈

涉及學段	文獻數量（篇）
幼稚園	1
幼稚園 - 小學	10
幼稚園 - 初中	1
幼稚園 - 高中	4
小學	57
小學 - 初中	12
小學 - 高中	28
初中	26
初中 - 高中	8
高中	40

從研究學科看（見表 5），K-12 教育文本難度的研究主要聚焦在閱讀（N=66）、科學（N=47）、英語二語（N=37）、社會學（N=35）等學科。此外，教科書是研究的主要文本類型（N=116）佔比 62%，其次是課外讀物（N=56），佔比 30%，對教輔資料（課程標準、試卷等）的研究較少，僅有 15 篇，佔比 8%。在目標語言方面，英語是最受關注的語言（N=139），佔比 74.3%。整體而言，K-12 教育文本難度的研究主要集中於對教科書的討論，尤其是英語類教科書（含英語二語學習），與之前對關鍵詞的突現分析結果吻合（見表 3）。這可能是因為大部分研究主要來源於英語國家和地區，且關於英語類文本難度的研究相對成熟。例如，關於文本難度量化評估的可讀性公式最早可追溯到 1921 年桑代克（Thorndike, 1921）編製的英語常用詞詞表。此外，隨着計算機技術和演算法的發展，研究者們構建了大型的、涵蓋多個年齡層的英語語料庫、文本特徵數據庫和可自動計算文本可讀性的軟件平台（如 Coh-Metrix、TextEvaluator），為英語文本難度的相關

研究提供了極大的便利和支持。類似的資源與平台在非英語中相對較少，這可能在很大程度上限制了研究者對非英語教材文本難度的研究。

表 5：研究文獻的學科分佈

學科	篇數（篇）	學科	篇數（篇）
閱讀（Reading）	66	俄語（Russian）	2
科學（Science）	47	希臘語（Greek）	1
英語二語（EFL）	37	波斯語（Persian）	1
社會學（Sociology）	35	泰語（Thai）	1
英語語言藝術（ELA）	8	阿拉伯語（Arabic）	2
數學（Math）	7	法語（French）	1
漢語（Chinese）	3	越南語（Vietnamese）	1

注：為減少誤差，統計時盡可能保留原文課程名稱，僅將原文中的物理、化學、生物和天文學歸為“科學”；將歷史、地理、經濟學、人類學、心理學和政治學歸為“社會學”。其中，閱讀學科包含英語母語閱讀 52 篇，非英語閱讀 14 篇：西班牙語 1 篇、法語 2 篇、土耳其語 3 篇、葡萄牙語 1 篇、荷蘭語 1 篇、希伯來語 1 篇、南非語 1 篇、孟加拉語 2 篇、波斯語 1 篇和瑞典語 1 篇。

（二）文本難度評估方法

關於文本難度的評估方法，美國共同核心國家標準（Common Core State Standards）提出從定量、定性、讀者與任務三個維度來評估文本難度（Hiebert, 2013）。定量維度指文本難度由可讀性公式來評估，涉及單詞長度、頻率和句子長度等量化指標；定性維度指文本難度靠專家評定，涉及文本意義、文本主題和層次等質性因素；讀者與任務維度指評估文本難度時考慮特定讀者（如知識和經驗）和特定任務（如任務要求和問題類型）的特點。本研究依據此分類框架，對納入文獻的文本難度評估方法進行統計分析。結果顯示，K-12 教育中文本難度的評估方法以可讀性公式的定量評估（N=156）為主，佔比 83%，其次為專家評估的定性分析框架評估（N=20），佔比 11%，而讀者與任務方面評估較少，僅有 15 篇（讀者特點），佔比 8%，其中 4 篇既使用了讀者特點評估，也使用了可讀性公式的定量評估。

1. 可讀性公式評估

文本難度的定量評估主要指運用不同類型的可讀性公式對文本進行難度估計，從研究發展上可分為傳統可讀性公式、基於認知理論的可讀性公式和基於統計語言模型的可讀性公式（Benjamin, 2012）。傳統可讀性公式注重基於文本特徵，如詞長、句長和詞頻等來評估文本難度；基於認知理論的可讀性公式則綜合考慮讀者特點對文本難度的影響，注重文本與讀者之間的匹配；基於統計語言模型的可讀性公式指利用自然語言處理技術和機器學習演算法來自動化計算文本特徵並評估文本難度，同時關注讀者與文本的交互影響，強調讀者在閱讀文本時的主觀能動性（Benjamin, 2012）。本研究統計了全部文獻（N=187）中使用的可讀性公式，考慮到部分研究綜合使用多種公式來評估文本難度，故統計結果為各公式在研究中出現的次數。如表 6 顯示，納入的全部文獻中有 84 篇研究使用了傳統可讀性公式，如 Fry、Flesch（Begeny & Greene, 2014）；有 43 篇研究使用了基於認知理論的可讀性公式，如 Lexile、Degrees of Reading Power（Roux et al., 2018）；有 49 篇研究使用基於統計語言模型的可讀性公式，如 Coh-Metrix、Text Evaluator（Liceralde et al., 2022）。

表 6：K-12 教育文本難度的定量評估

可讀性公式分類	核心關注點	可讀性公式列舉
傳統可讀性公式（84）	文本語言特徵（詞頻、句長）	Fry、Flesch
認知理論可讀性公式（43）	讀者特點：文本與讀者間匹配	Lexile、DRP
統計語言模型可讀性公式（49）	機器學習和自然語言處理	Coh-Metrix、TextEvaluator

注：括弧中的數字指作者編碼次數，如「傳統可讀性公式（84）」指傳統可讀性公式在 84 篇文獻中用於評估文本難度。

此外，統計分析了可讀性公式在不同學習階段的使用情況。由於部分研究同時涵蓋了不同年級，統計時計算了各年級出現的次數。資料顯示，可讀性公式主要用於評估小學和中學階段小學，N=88；初

中，N=67；高中，N=69）的文本難度，而較少在幼稚園階段（N=5）的文本難度評估中使用。這可能與幼兒階段文本本身的特點有關。

2. 專家評估

文本難度的定性評估主要依靠專家基於定性分析工具指南對文本進行難度評定。本研究依據 Person et al.（2014）對文本難度定性框架的分類，對 187 篇納入文獻中定性分析工具的具體使用情況進行了統計（見表 7）。僅有 20 篇研究使用了定性分析框架來評估文本難度。其中，6 篇使用文本分級系統（5 篇 RR books，1 篇 STAS-1）（Pitcher & Fang, 2007；Hoffman et al., 2001），13 篇使用評分標準範例系統（CCSS）（Sierschynski et al., 2014），1 篇使用文本地圖系統（Text Maps）（Chambliss, 1994）。整體而言，定性分析中使用 CCSS 評估文本難度的研究最多，這可能與其本身政策性特點有關。³

表 7：K-12 教育文本難度的定性評估

定性分析框架	評估工具例舉	具體分析框架
文本分級系統（6） （Text-leveling systems）	RR books（5） （Reading Recovery）	書籍和印刷品特徵 內容、主題和思想 文本結構 語言和文學元素
	STAS-1（1）（Scale for Text Accessibility and Support）	可解碼性：三點量表 可預測性：五點量表
評分標準範例系統（13） （Rubric and exemplar systems）	CCSS（13） （Common Core State Standards）	文本意義、層次 文本結構 語言規範、清晰 知識需求
文本地圖系統（1） （Text-mapping systems）	Text Maps（1）	圖形分類法 故事語法流程圖

注：表中內容的整理依據 Person et al.（2014）提出的文本難度定性框架分類。括弧中的數字指作者編碼次數，如「文本分級系統（6）」指文本分級系統在 6 篇文獻中用於評估文本難度。

3 美國《共同核心國家標準》（Common Core State Standards, CCSS），由全美州長協會（NGA）和州立學校官員委員會（CCSSO）於 2010 年聯合制定，最初 48 個州共同參與，旨在解決各州教育標準差異過大的問題。截至 2024 年，41 個州及哥倫比亞特區仍在採用。CCSS 主要涵蓋英語語言藝術（ELA）和數學領域。在英語語言藝術方面，着重對 K-12 學生分析複雜文本及閱讀能力提出要求，並構建了包含定量指標、定性分析及讀者任務情境的文本難度評估框架。

此外，本研究亦統計了定性框架在不同學段中的使用情況（次數）：幼稚園（N=4）、小學（N=15）、初中（N=5）、高中（N=2），發現定性分析主要聚焦於幼稚園和小學低年級階段（一年級至三年級），這可能與低幼階段（尤其是幼稚園階段）文本本身的特點有關。結合研究內容看，定性分析框架主要用於評估幼稚園和小學階段圖畫書的文本難度，討論圖書分級系統對低幼階段文本難度評估的有效性以及分析定性評估對小學教師閱讀教學中文本篩選的影響（Dzaldov & Peterson, 2005）。

3. 讀者與任務評估

從非文本維度（讀者、任務）對文本難度進行評估的研究較少（N=15），已有研究均基於讀者因素（N=15），未發現有研究基於任務因素。其中，7 項研究使用不同級別的量表（如 5 點量表，非常容易到非常難；3 點量表，由低到高），由學生作為讀者來對文本難度進行主觀評定（Pishghadam & Abbasnejad, 2016）；有 8 項研究則根據文本編製閱讀理解測驗，由學生作為讀者在閱讀理解測驗上的成績來標定相應文本的難度（Gecit, 2010）。

（三）文本難度影響因素

文本任務反應理論（Text-Task-Respondent Theory，TTR）認為，個體對文本的理解不僅受文本特徵的影響，而且受任務和讀者特點的影響（Hoover & Gough, 1990）。文本特徵指文本本身的特點，如詞彙難度、句子複雜度和句法難度等。讀者特點則指讀者在興趣、閱讀動機、背景知識、經驗技能等方面的個體差異。任務特點指讀者從文本中獲得資訊或完成特定任務所需的要求，如閱讀目的和問題類型。例如，有研究探討文本詞彙難度（文本特徵）對文本難度的影響（White, 2012）；也有研究分析個體認知技能和先驗知識（讀者因素）對文本難度的影響，發現個體的單詞閱讀能力、詞彙知識和推理理解能力、主題知識等均影響其對文本的理解（文本難度）（McNamara et al., 2011）。另有研究探討問題類型和任務要求（任務因素）對文本難度

的影響，如通過比較讀者在不同問題類型上的閱讀理解表現，發現推理型和批判分析型問題均比直接陳述型問題更具挑戰性，能提高文本難度（Spencer et al., 2019）；此外，閱讀課程認知要求（如問題數量、回答問題所需的信息量和問題類型的豐富程度等）或教師教學實踐要求（如任務解釋的清晰度和要求學生完成的程度等）均能增加或降低讀者理解文本的難度（Stevens et al., 2015；Viriot-Goeldel & Crinon, 2016）。

本研究基於 TTR 理論，對在研究結果中明確指出影響文本難度因素的 187 篇納入文獻進行分類編碼與分析。

通過三個層級的歸納和總結，本研究提煉出影響文本難度的核心因素（見表 8）。例如，將三級維度如「詞彙、句子、語法」等歸納為二級維度「文本語言特徵」，並進一步將二級維度如「文本語言特徵、文本形式、文本內容」歸納為一級維度「文本特徵」。由於大量研究綜合考慮了文本難度的多個影響因素，所以高一級維度的編碼數量並非低一級維度編碼數量的簡單加和。例如，一級維度「文本特徵」因素在 171 篇文獻中被發現影響文本難度，包括二級維度的文本語言特徵（N=166）、文本形式（N=79）和文本內容（N=34），而一級維度的編碼數量（N=171）遠少於各二級維度編碼數量的總和（N=279）。這說明文本難度並非受單一特徵影響，而是多種因素綜合作用的結果。另外，與讀者（N=6）和任務（N=5）特點相比，文本特徵（N=171）的研究更多，這可能是因為可讀性公式及其軟件應用的發展為相關研究提供了便利。相比之下，讀者和任務特點的研究則更為複雜，涉及多種因素，未來可加強此方面的研究。

表 8：K-12 教育文本難度的影響因素

一級維度	二級維度	三級維度
文本特徵 (171)	文本語言特徵 (166)	詞彙 (161)、句子 (92)、語法 (57) 語義 (57)、文本風格 (3)
	文本形式 (79)	篇章結構 (10)、文本長度 (5) 體裁 (37)、文本凝聚力 (39)
	文本內容 (34)	文本主題 (10)、資訊密度 (23) 文本意義 (13)、文本目的 (13)
讀者特點 (6)	特點表現 (6)	主題知識 (4)、認知技能 (2)
任務特點 (5)	特點表現 (5)	問題類型 (3)、任務要求 (2)

注：括弧中的數字指作者編碼次數，如「詞彙 (161)」指詞彙在 161 篇文章中被發現影響文本難度。

(四) 文本難度支持策略

值得關注的是，從教育的角度，在說明學生理解複雜文本時，任務（支持策略）因素相較於於文本和讀者因素更可控。已有的文本難度評估方法和影響因素的深入研究也為支持策略的研究提供了基礎。例如，教師可通過使用合適的評估方法來篩選文本，靈活調整給學生的教學材料和閱讀指導；研究者可根據文本難度影響因素開展干預研究，探索適用於不同學生群體的閱讀支持策略。然而，本研究中僅有 8 篇納入文獻對 K-12 教育背景下的文本難度支持策略進行了研究。在教學策略方面，有研究採用隨機干預實驗（N=1）對比知 - 想 - 學（Know-Want-Learn，K-W-L）策略與聽 - 讀 - 討論（Listen-Read-Discuss，L-R-D）策略對學生文本理解的影響（Lupo et al., 2019），課後學生閱讀理解測試結果表明 K-W-L 能夠有效幫助學生閱讀複雜文本；另 5 篇研究通過類似實驗範式（N=5），表明互動式漸進提示 / 支架（Scaffolding）策略能有效支持學生閱讀複雜文本（Reynolds, 2021）；亦有研究（N=1）通過對比同一組學生在干預前後的閱讀理解能力，表明複雜文本分析教學法（Complex Text Analysis）能有效幫助一年級學生閱讀複雜文本（Witte, 2016）。在閱讀策略方面（N=1），研究者對比不同閱讀情境下（細讀干預和非干預）學生的閱讀成績，

發現細讀（Close Reading）策略能有效支持學生閱讀複雜文本（Fisher & Frey, 2014）。整體而言，與教學策略研究相比，閱讀策略方面的研究較少。

四、研究結論與建議

（一）研究結論

本研究從研究特徵、評估方法、影響因素和支持策略四方面對 K-12 教育背景下的文本難度研究進行系統性綜述，得出以下研究結論。在研究特徵方面，既往研究主要聚焦於文本語言複雜性、英語教材難度及基於大型語料庫來分析文本難度；研究學段集中在小學至高中階段，少有對幼稚園階段文本難度的分析；研究材料以英語類教材為主，對非英語類文本的關注有限。在評估方法方面，採用可讀性公式的定量評估是評定文本難度的主要方法，尤其是在小學和中學階段。在影響因素方面，不同類型的文本特徵是影響文本難度的主要因素，讀者和任務因素也發揮了一定作用。在支持策略方面，雖然已有研究提出了互動式支架、複雜文本分析法等策略，但整體而言較少研究複雜文本的閱讀支持策略。

（二）研究建議

根據上述研究結論，本研究提出以下三方面關於 K-12 階段文本難度未來研究的相關建議，供相關學者及教育工作者參考。

1. 注重幼兒階段的文本難度研究

本研究發現，國內外已有的文本難度研究主要集中在小學和中學階段，對幼稚園階段的文本關注較少。另外，本研究也發現幼兒階段文本難度的評估方法主要以定性分析框架為主，而非定量的可讀性公式。這與可讀性公式特點以及幼兒階段文本本身的特點有關。可讀性公式的本質是建立文本特徵和文本難度之間預測關係的統計模型，主要通過對可量化的文本特徵，如句子長度、單詞頻率或句法複雜

性，進行計算，以評估文本難度。此類測量方法局限於可量化的文本因素，較難納入可用於評估文本難度的質性因素，如文章主題、意義層次、圖畫特徵等，故類似可讀性公式的定量評估方法難以相容幼兒階段文本。與其他階段相比，幼兒階段的閱讀文本多為繪本，篇幅較短、圖畫居多、書面文字較少，能夠量化的文本因素比重少，且更多地結合了插圖、設計和排版等多模態特徵（Kachorsky, 2017）。相較而言，注重從主題、意義層次和圖像等維度的定性分析更適合評估幼兒階段的文本難度。然而，目前國內外針對幼兒階段的文本難度研究較少，導致教師或家長在早期閱讀活動中對文本的選擇仍依賴直覺與經驗，具較強的主觀性，而非基於理論與實證。因此，未來學界有必要加強對幼兒階段文本難度的研究，豐富已有的閱讀學習理論與實證研究，指導一線教學與家庭閱讀活動，為幼兒提供適合其早期閱讀能力萌芽的文本，也為早期閱讀活動的開展和閱讀材料的設計提供參考。

2. 注重文本語言多樣性，研究非英語類文本的難度

基於現有文獻分析，本研究發現，當前文本難度研究主要圍繞英語語言特徵展開，有關評估方法、影響因素及支持策略的探討也多以英語教材為樣本，對其他語言文本的研究較少。不同語言因其自身獨特的語法結構、詞彙特點和語言風格等，故無法將英文可讀性公式直接應用於其他語言文本的難度評估。雖有研究已開始關注非英語類文本的難度，如探索阿拉伯語文本可讀性公式開發（Ghani et al., 2014），分析土耳其語教科書的可讀性特徵（Turkben, 2019），但整體研究仍較少。為更好地滿足多語言環境下的教育需求，未來的研究需關注文本語言的多樣性，探討不同語言的語言特徵、詞彙使用和句法複雜性對文本難度影響，開展跨語言的文本難度研究，為不同語言背景下的學習者（讀者）、不同的閱讀目的（任務）提供適合的文本（文本）。此外，本研究發現，已有研究的文本材料類型以 K-12 階段（幼稚園、初中和高中）各科教材為主，未來需加強對不同年級或年齡段課外讀物難度的研究，包括開發評估課外讀物文本難度的平台，探索各年齡段的分級閱讀體系等。

3. 注重文本難度支持策略方面的研究

儘管已有研究提出了互動式支架和細讀等閱讀支持策略來幫助學習者閱讀複雜文本 (Reynolds & Goodwin, 2016)，但國內外此類研究整體仍不足。鑒於 K-12 階段培養學生閱讀技能的重要性，有必要加強此領域的研究。未來的研究可充分借鑒文本難度評估方法和影響因素方面的研究成果，從不同角度，如年級、讀者特點、活動類型和學習場景等方面展開研究。例如，比較不同年級或年齡層學生對文本難度的感知和理解能力差異。通過分析幼兒、小學生、初中生和高中生在閱讀不同難度文本時所面臨的挑戰和困難 (Amendum et al., 2016)，及其對閱讀支持策略的需求和反應，來探索針對性的閱讀策略或教學策略。同時，關注學生（讀者）的個體差異，如語言水準、閱讀技能、認知能力和學習風格等，為不同類型的學生提供個性化的支持策略，包括有閱讀障礙或特殊學習需求的學生 (Venable, 2003)。另外，可分析不同類型的閱讀活動（如課堂教學、小組合作和個人獨立閱讀等）對文本難度的影響 (Valencia et al., 2014)，研究適應不同活動類型的支持策略。此外，關注不同學習場景下的文本難度支持策略。例如，在學校課堂、家庭環境和線上學習平台等不同場景中，探索相應的文本難度支持策略，以提供更具針對性的支持。未來也需開展跨學科的研究，將文本難度支持策略研究與其他領域研究結合，如閱讀心理學、教育技術和學習障礙研究等 (Atanga et al., 2020)，為文本難度支持策略的開發和實施提供更全面和多維度的思考視角。

參考文獻

- 邊偉、王曉誠、車麗娜 (2017)：基於學生視角的語文教科書資訊文本難度評析，《教育科學研究》，(10)，48-53。
- 柏曉鵬、吉伶俐 (2022)：篇章結構特徵對文本可讀性的影響，《語言文字應用》，(3)，62-72。
- 程勇、徐德寬、董軍 (2020)：基於語文教材語料庫的文本閱讀難度分級關鍵因素分析與易讀性公式研究，《語言文字應用》，(2)，132-143。
- 程勇、徐德寬、董軍 (2020)：基於多元語言特徵與深度特徵融合的中文文本閱讀難度自動分級研究，《中文資訊學報》，(4)，101-110。
- 程勇、徐德寬、呂學強 (2019)：基於多元特徵的文本閱讀難度自動分級研究，《數據分析與知識發現》，(7)，106-112。
- 李文彪、吳雲芳 (2023)：基於神經網路模型的漢語文本難度分級，《中文資訊學報》，(2)，158-168。
- 劉苗苗、李燕、王欣萌、甘琳琳、李虹 (2021)：分級閱讀初探：基於小學教材的漢語可讀性公式研究，《語言文字應用》，(2)，116-126。
- 劉瀟 (2015)：文本易讀度相關研究述評，《湖北大學學報·哲學社會科學版》，(5)，141-146。
- 吳思遠、於東、江新 (2020)：漢語文本可讀性特徵體系構建和效度驗證，《世界漢語教學》，(1)，81-97。
- 王蕾 (2022)：文本可讀性公式研究發展階段及特點，《語言教學與研究》，(2)，29-40。
- 吳思遠、蔡建永、於東 (2018)：文本可讀性的自動分析研究綜述，《中文資訊學報》，(12)，1-10。
- 張秋玲、牛青森、趙甯寧 (2022)：中學語文教科書文言選文難易度評量模型檢驗，《語言文字應用》，(3)，49-60。
- Alexander, P. A., & Jetton, T. L. (2000). *Learning from text: A multidimensional and developmental perspective*. Mahwah, NJ: Erlbaum.
- Amendum, S. J., Conradi, K., & Liebfreund, M. D. (2016). The push for more challenging texts: An analysis of early readers' rate, accuracy, and comprehension. *Reading Psychology*, 37(4), 570-600. <https://doi.org/10.1080/02702711.2015.1072609>
- Atanga, C., Jones, B. A., Krueger, L. E., & Lu, S. (2020). Teachers of students with learning disabilities: Assistive technology knowledge, perceptions, interests, and barriers. *Journal of Special Education Technology*, 35(4), 236-248. <https://doi.org/10.1177/0162643419864858>
- Balyan, R., McCarthy, K. S., & McNamara, D. S. (2020). Applying natural language processing and hierarchical machine learning approaches to text difficulty classification. *International Journal of Artificial Intelligence in Education*, 30(3), 337-370. <https://doi.org/10.1007/s40593-020-00201-7>
- Benjamin, R. G. (2012). Reconstructing readability: Recent developments and recommendations in the analysis of text difficulty. *Educational Psychology Review*, 24, 63-88. <https://www.jstor.org/stable/43546780>
- Begeny, J. C., & Greene, D. J. (2014). Can readability formulas be used to successfully gauge difficulty of reading materials?. *Psychology in the Schools*, 51(2), 198-215. <https://doi.org/10.1002/pits.21740>
- Bomer, R., & Maloch, B. (2011). Research and Policy: Relating Policy to Research and Practice: The Common Core Standards. *Language Arts*, 89(1), 38-43. <https://doi.org/10.58680/la201117158>
- Chall, J. S. (1983). *Stages of reading development*. McGraw-Hill.
- Chambliss, M. J. (1994). Evaluating the quality of textbooks for diverse learners. *Remedial and special education*, 15(6), 348-362. <https://doi.org/10.1177/074193259401500604>
- Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for information Science and*

- Technology*, 57(3), 359-377.
<https://doi.org/10.1002/asi.20317>
- Dzaldov, B. S., & Peterson, S. (2005). Book leveling and readers. *The reading teacher*, 59(3), 222-229.
<https://doi.org/10.1598/RT.59.3.2>
- Fisher, D., & Frey, N. (2014). Close reading as an intervention for struggling middle school readers. *Journal of adolescent & Adult literacy*, 57(5), 367-376. <https://doi.org/10.1002/jaal.266>
- Fitzgerald, J., Elmore, J., Koons, H., Hiebert, E. H., Bowen, K., Sanford-Moore, E. E., & Stenner, A. J. (2015). Important text characteristics for early-grades text complexity. *Journal of Educational Psychology*, 107(1), 4. <https://doi.org/10.1037/a0037289>
- Ghani, K. A., Noh, A. S., & Yusoff, N. M. (2014). Linguistic features for development of Arabic text readability formula in Malaysia: A preliminary study. *Middle-East Journal of Scientific Research*, 19(3), 319-331. 10.5829/idosi.mejsr.2014.19.3.13597
- Gamson, D. A., Lu, X., & Eckert, S. A. (2013). Challenging the research base of the Common Core State Standards: A historical reanalysis of text complexity. *Educational Researcher*, 42(7), 381-391. <https://doi.org/10.3102/0013189X13505684>
- Gecit, Y. (2010). The Evaluation of High School Geography 9 and High School Geography 11 Text Books With Some Formulas of Readability. *Educational Sciences: Theory and Practice*, 10(4), 2205-2220.
- Hiebert, E. H. (2013). Supporting students' movement up the staircase of text complexity. *The Reading Teacher*, 66(6), 459-468. <https://doi.org/10.1002/TRTR.1149>
- Hoffman, J. V., Roser, N. L., Salas, R., Patterson, E., & Pennington, J. (2001). Text leveling and "little books" in first-grade reading. *Journal of Literacy Research*, 33(3), 507-528. <https://doi.org/10.1080/10862960109548121>
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and writing*, 2, 127-160. 10.1007/BF00401799
- Harlaar, N., Dale, P. S., & Plomin, R. (2007). From learning to read to reading to learn: Substantial and stable genetic influence. *Child Development*, 78(1), 116-131. <https://doi.org/10.1111/j.1467-8624.2007.00988.x>
- Kachorsky, D., Moses, L., Serafini, F., & Hoelting, M. (2017). Meaning making with picturebooks: Young children's use of semiotic resources. *Literacy Research and Instruction*, 56(3), 231-249. <https://doi.org/10.1080/19388071.2017.1304595>
- Licalalde, V. R. T., Loukina, A., Beigman Klebanov, B., & Lockwood, J. R. (2022). Beyond text complexity: Production-related sources of text-based variability in oral reading fluency. *Journal of Educational Psychology*, 114(1), 16.
<https://doi.org/10.1037/edu0000532>
- Lupo, S. M., Tortorelli, L., Invernizzi, M., Ryoo, J. H., & Strong, J. Z. (2019). An exploration of text difficulty and knowledge support on adolescents' comprehension. *Reading Research Quarterly*, 54(4), 457-479.
<https://doi.org/10.1002/rrq.247>
- Mesmer, H. A., Cunningham, J. W., & Hiebert, E. H. (2012). Toward a theoretical model of text complexity for the early grades: Learning from the past, anticipating the future. *Reading research quarterly*, 47(3), 235-258. <https://doi.org/10.1002/rrq.019>
- Mikk, J., & Elts, J. (1999). A reading comprehension formula of reader and text characteristics. *Journal of Quantitative Linguistics*, 6(3), 214-221. 10.1076/jqul.6.3.214.6158
- McNamara, D. S., Ozuru, Y., & Floyd, R. G. (2011). Comprehension challenges in the fourth grade: The roles of text cohesion, text genre, and readers' prior knowledge. *International electronic*

- journal of elementary education*, 4(1), 229-257.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *bmj*, 372.
<https://doi.org/10.1136/bmj.n71>
- Pitcher, B., & Fang, Z. (2007). Can we trust levelled texts? An examination of their reliability and quality from a linguistic perspective. *Literacy*, 41(1), 43-51. <https://doi.org/10.1111/j.1467-9345.2007.00454.x>
- Pearson, P. D., & Hiebert, E. H. (2014). The state of the field: Qualitative analyses of text complexity. *The Elementary School Journal*, 115(2), 161-183.
<https://doi.org/10.1086/678297>
- Pishghadam, R., & Abbasnejad, H. (2016). Emotioncy: A potential measure of readability. *International Electronic Journal of Elementary Education*, 9(1), 109-123.
- Roux, K., Mtsatse, N., & Long, C. (2018). A comparison of text difficulty in systemic assessment using Lexile theory. *Southern African Linguistics and Applied Language Studies*, 36(4), 291-300. 10.2989/16073614.2018.1548290
- Reynolds, D. (2021). Scaffolding the academic language of complex text: An intervention for late secondary students. *Journal of Research in Reading*, 44(3), 508-528. <https://doi.org/10.1111/1467-9817.12353>
- Reynolds, D., & Goodwin, A. (2016). Supporting students reading complex texts: Evidence for motivational scaffolding. *Aera Open*, 2(4). <https://doi.org/10.1177/2332858416680353>
- Spencer, M., Gilmour, A. F., Miller, A. C., Emerson, A. M., Saha, N. M., & Cutting, L. E. (2019). Understanding the influence of text complexity and question type on reading outcomes. *Reading and writing*, 32, 603-637.
 10.1007/s11145-018-9883-0
- Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Rand Corporation.
- Siddaway, A. P., Wood, A. M., & Hedges, L. V. (2019). How to do a systematic review: a best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual review of psychology*, 70(1), 747-770. <https://doi.org/10.1146/annurev-psych-010418-102803>
- Sierschynski, J., Louie, B., & Pughe, B. (2014). Complexity in picture books. *The Reading Teacher*, 68(4), 287-295. <https://doi.org/10.1002/trtr.1293>
- Stevens, R. J., Lu, X., Baker, D. P., Ray, M. N., Eckert, S. A., & Gamson, D. A. (2015). Assessing the cognitive demands of a century of reading curricula: An analysis of reading text and comprehension tasks from 1910 to 2000. *American Educational Research Journal*, 52(3), 582-617. <https://doi.org/10.3102/0002831215573531>
- Thorndike, E. L. (1921). *The teacher's workbook*. New York, NY: Teacher's College, Columbia University.
- Turkben, T. (2019). Readability Characteristics of Texts in Middle School Turkish Textbooks. *Educational Policy Analysis and Strategic Research*, 14(3), 80-105. 10.29329/epasr.2019.208.5
- Viriot-Goedel, C., & Crinon, J. (2016). The use of complex reading materials in first grade. *L1-Educational Studies in Language and Literature*, 16, 1-21. <https://doi.org/10.17239/L1ESLL-2016.16.01.02>
- Venable, G. P. (2003). Confronting complex text: Readability lessons from students with language learning disabilities. *Topics in Language Disorders*, 23(3), 225-240. 10.1097/00011363-200307000-00008

- Valencia, S. W., Wixson, K. K., & Pearson, P. D. (2014). Putting text complexity in context: Refocusing on comprehension of complex text. *The Elementary School Journal*, 115(2), 270-289. <https://doi.org/10.1086/678296>
- White, M. D., & Marsh, E. E. (2006). Content analysis: A flexible methodology. *Library trends*, 55(1), 22-45. 10.1353/lib.2006.0053
- White, S. (2012). Mining the text: 34 text features that can ease or obstruct text comprehension and use. *Literacy Research and Instruction*, 51(2), 143-164. <https://doi.org/10.1080/19388071.2011.553023>
- Witte, P. G. (2016). Teaching first graders to comprehend complex texts through read - alouds. *The reading teacher*, 70(1), 29-38. <https://doi.org/10.1002/trtr.1467>
- Yang, S. J. (1971). *A readability formula for Chinese language*. The University of Wisconsin-Madison.

A Systematic Review of Research on Text Difficulty in the Context of K-12 Education: From 1948 to 2024

ZHANG, Xingli LI, Yixun QIU, Zhimin

Abstract

What makes one text more difficult than another has been the interest of researchers and educators over the past century. To guide theoretical development and research directions, this systematic review sought to identify (1) the emerging research themes, (2) the methods (qualitative, quantitative, and mixed) to measure text complexity, and (3) the key findings in existing text difficulty research. We targeted empirical research on text complexity in the context of K-12 education that was conducted outside of China and published in peer-reviewed English journals. Based on these predefined inclusion criteria, we conducted a comprehensive search in the major electronic databases (e.g., EBSCO and Web of Science) and the four previous reviews (e.g., Magdalena, 2021). A total of 11,610 articles were found between 1,948 (the earliest time we could find records) and April 2024, with 10,270 articles after removing duplicates. We screened these records by titles and abstracts (592 articles remained) and then by full texts, leading to 187 articles in the final pool. Each article was coded, and the findings were synthesized. Our key findings include (1) most articles (78.0%) were published after 2010, and primarily focused on the complexity of English texts (74.3%), with a recent proliferation in other languages (e.g., Russian, Arabic, Thai, Turkish, and German); (2) more than half of the studies developed quantitative-based readability formulas (83.0%), some adopted qualitative schemes (11.0%), and reader ratings were rarely used (8.0%); (3) most studies focused on primary school and above (91.4%), and less on

kindergarten (8.6%); (4) Little existing work has focused how to support students in comprehending complex texts. Future research on texts for young children using a mixed approach is warranted to support reading acquisition in a wider age range.

Keywords: K-12; Text Difficulty, Systematic Review, Empirical Research, Content Analysis

ZHANG, Xingli, Department of Early Childhood Education, The Education University of Hong Kong, HK.

* LI ,Yixun, Department of Early Childhood Education, The Education University of Hong Kong, HK.
(corresponding author)

QIU , Zhimin, Department of Early Childhood Education, The Education University of Hong Kong, HK.