education sciences

MDPI

Review

The Chatbots' Challenge to Education: Disruption or Destruction?

Menucha Birenbaum

School of Education, Tel Aviv University, Tel Aviv 69978, Israel; biren@tauex.tau.ac.il

Abstract: The article addresses the positive and negative implications of the growing spread of chatbots based on large language models (LLMs) on instruction, learning, and assessment in education. It is based on extensive conversations with ChatGPT regarding pedagogy-related issues and relevant documents. Discussed are the challenges of chatbots like ChatGPT to educators—on the one hand, their potential to advance deep learning and the roles of the instructor and the school context in causing it to happen. On the other hand, it underscores the pedagogical drawbacks of improper usage of such chatbots and the instructional practices and school contexts that could escalate learning. Three school-culture components, namely classroom learning, teacher professional learning, and school leadership, are the essential aspects of pedagogical approaches that, in a particular constellation, could enhance and, in another, impede a chatbot's potential to advance deep learning. The underlying theoretical framework is assessment-driven, contrasting assessment for learning (AfL) and assessment for grading, distinguishing assessment cultures from testing cultures. Patterns of chatbot usage that align with the principles of each culture are discussed. A sample of quotes from the conversations with ChatGPT is presented to support the insights gained from the chatting experience and the conclusions drawn.

Keywords: LLMs; ChatGPT; assessment for learning (AfL); improving classroom teaching and learning

check for updates

Citation: Birenbaum, M. The Chatbots' Challenge to Education: Disruption or Destruction? *Educ. Sci.* **2023**, *13*, 711. https://doi.org/10.3390/educsci13070711

Academic Editor: Diego Vergara

Received: 19 May 2023
Revised: 27 June 2023
Accepted: 11 July 2023
Published: 13 July 2023

Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

A shift from analytic to generative AI has marked the field of AI in recent years. Prominent in this direction are large language models (LLMs); their accelerated widespread use raises many questions regarding their utility. Much has been written about the advantages and disadvantages of those generative systems over several years of usage in fields such as health care [1], journalism [2], retail [3], the finance industry [4], and other industries.

An LLM that, as of early 2021, was the largest neural network is OpenAI's GPT-3 (which stands for third-generation Generative Pre-trained Transformer). It is a deep learning neural network model (with over 175 billion machine learning parameters) trained using internet data to generate realistic human texts of various types. GPT-3 uses natural language generation and processing to understand and generate natural human language text. It can create anything with a text structure, and has been used to create articles, stories, poetry, music, jokes, comic strips, news reports, test and quiz items, assessment rubrics, provide feedback, and to generate summarization and programming codes [5].

Teachers utilize chatbots like GPT-3 to design tutoring interventions and assessment tools customized to students' different preferences and levels, give feedback on students' performances, and stimulate students' curiosity and higher-order thinking [6]. Students often use GPT-3 to retrieve and summarize the information needed to perform their learning and assessment tasks, including research projects and essays. However, some expect it to write the papers for them [7].

The recent release of ChatGPT (on 30 November 2022), soon to be incorporated into several Microsoft applications, and the rushed release of Bard, supported by Google, have stimulated heated discussions among educators about the potential impact of generative AI LLMs (hereafter, chatbots) on learning.

The current article focuses on using such tools in education, particularly ChatGPT (a variant of the GPT-3 model optimized for human dialogue). It crossed one million users within one week of its launch [8]. Addressed is an aspect that has not yet been explored in-depth regarding chatbots, namely, their impact on significant learning. Or, in other words, what are the opportunities such tools present for fostering learning that could lead to transforming education? In contrast, what are the barriers or setbacks that, by using such tools, could impede significant learning?

A SWOT analysis [9] on educational usage of chatbots, like GPT-3 (including its variant ChatGPT), marks the following strengths, weaknesses, opportunities, and threats:

Strengths: Chatbots are accessible at any time, any place, have a friendly interface, are less expensive than human tutors, can be tailored to students' preferences and needs, and help detect plagiarism (for instance, via AI text classifiers) and other academic dishonesty acts [5]. Moreover, ChatGPT is grammatically fluent, setting an example of a well-structured verbal response; it can also generate programming code, find bugs in existing code, translate between programming languages, and even help cyber security teams [10]. The following three capabilities of ChatGPT are listed on its sign-on screen: *"Remembers what user said earlier in the conversation; allows user to provide follow-up corrections, and trained to decline inappropriate requests"*.

Overall, chatbots' significance is highly acknowledged. In his blog, Bill Gates wrote that recent AI developments, like ChatGPT, are as revolutionary as mobile phones and the internet [11] (It should be noted that Bill Gates was involved in the development of ChatGPT as a philanthropist and an advisor).

Weaknesses: Chatbots can only do what they were trained to and refer to information from databases to which they were exposed; hence, they may not be able to answer complex questions. Other limitations of chatbots are their inability to understand and respond to users' emotions and feelings, and their dependency on technology infrastructure, which could affect their reliability. Moreover, they are known to lose coherence over long text passages and fabricate information (a characteristic often called *artificial hallucination*) [12,13]. Another limitation of chatbots concerns the security of the users' data [6].

To the question of whether GPT-3, like other chatbots, constitutes a good model of human language, Noam Chomsky responded negatively, arguing that GPT-3 does not provide adequate explanations, but merely mimics statistical patterns of how language has been used in its database. Hence, its capacity is to predict sequences of words. He further asserts that GPT-3 cannot distinguish truth from fiction, and concludes that, unlike human minds who try to relate their language to what they know about the world, chatbots do not; their comprehension of the world is disorganized, bound to the input on which they were trained, and lacks genuine abstraction [14].

Another weakness relates to the adequacy of the information resources to which the chatbots were exposed during their training. For instance, ChatGPT was trained on open-access internet information, limiting its knowledge of scientific research published in paid subscription journals and books. Moreover, the cut-off date of its training was September 2021, meaning that it has not been systematically exposed to publications after that date [6].

Opportunities: If used competently, chatbots have the potential to open up a vast array of options for the sustainable transformation of education. (For instance, they can assist in transforming instructional tasks from the currently practiced tasks requiring information retrieval and summarization to tasks requiring higher-order skills, such as critical thinking.) Chatbots' usage can also increase students' curiosity, particularly if they are free to choose topics for the chats (for instance, in multi-user virtual environments integrated into modern classrooms) [15].

Gates [11] argues that computers failed to meet the expectation of changing education, maintaining that they have not had a meaningful effect on any measure of students' achievement. However, given the recent developments in AI, he claims that in the next five to ten years, AI-driven software will finally fulfill the promise to revolutionize how people teach and learn.

Threats: OpenAI has warned against the following risks associated with the use of GPT-3 in education: risks related to personalization (including violating student privacy, providing discriminatory treatment, and eliciting unhealthy habits); risks associated with failure in detecting academic dishonesty and plagiarism (referring to the inadequacy of the AI text classifier developed by the company to detect AI-generated content); and risks related to racial discrimination due to inherent biases in the training contents [6]. Consequently, the following warning note to ChatGPT's users pops up on the sign-on screen: *"While we have safeguards in place, the system may occasionally generate incorrect or misleading information and produce offensive or biased content. It is not intended to give advice . . . Please don't share any sensitive information in your conversations"*. These warnings add to growing concerns about future potential risks of such AI tools to society [16]. Hence, the urgent call of AI experts and industry executives for a six-month pause in the race to develop more powerful systems, a pause to be used for jointly developing safety protocols for advanced AI systems [17].

School Context

Research has marked significant contextual school effects on student outcomes [18,19], underscoring the school context as a substantial determinant in education reforms geared toward sustainable teaching and learning improvements [20].

Viewing schools through an assessment lens reveals a variety of school cultures ranging from testing cultures (TC) to assessment cultures (AC) [21]. The former focuses on making the grade by adhering to conventional teaching and assessment approaches. The latter focuses on learning, adopting an assessment for learning (AfL) pedagogy that makes learning visible to teachers and students, so that they can make improvements while it happens [22–25]. Furthermore, while TC schools adhere to external accountability demands [26], AC schools exhibit professional accountability that guides their practice [20,27].

Conceptualizing AC through a complexity lens [28] points to continuous reciprocal interactions within and between three systems nested in AC: classroom learning (hereafter, CL), teacher professional learning (hereafter, TPL), and school leadership (hereafter, SL). The interactions enhance the schools' internal coherence, which Elmore and his associates [29] perceive as *"a school's capacity to engage in deliberate improvements in instructional practice and student learning across classrooms over time, as evidenced by educator practices and organizational processes that connect and align work across the organization."* (p. 3). According to complexity principles, the more frequent and powerful the interactions among the agents, the more influence they are likely to have on the agents' behavior and, consequently, the more likely the emergence of new patterns; hence, the more likely the organization will be to experience renewal [30].

Three research questions were addressed in the current study:

1. What constellations of school culture components (CL, TPL, and SL) in which chatbots' usage patterns have the potential to promote deep learning, and why?
2. What constellation of school culture components in which chatbots' usage patterns can impede deep learning, and why?
3. What insights can be generated from the conversations with ChatGPT; in particular, what factors made some chats more fruitful than others?

2. Method

2.1. The Chats

The article is based on 50 chats (comprising 118 segments) with ChatGPT about pedagogical issues. More than one segment was included in 30 chats (ranging from 2 to 13 segments). Those chats consisted of follow-up questions and comments. The list of topics addressed in the chats, the content categories, and the segments per chat appear in Table A1 in Appendix A. Table A2 in Appendix A displays the classification scheme. Evident in Appendix A are also the various task types requested from ChatGTP.

The average user prompt per segment comprised 17.34 words (SD = 20.68) (ranging from 1 to 177 words). The average ChatGPT response per segment contained 215.36 words (SD = 128.51) (ranging from 64 to 1093). A total of 60% of the chats (30 chats) concerned the potential of school culture based on assessment for learning (AfL) to advance learning, accountability models, and the roles AI chatbots can play in supporting or impeding in-depth understanding and the causes for each. Another cluster of 19 chats concerned curiosity and the joy of learning, topics that current research and interventions aim to nurture in students, thus empowering them to become curious life-long learners. Finally, one chat concerned ChatGPT's learning. Data from pedagogy-related chats (school culture, accountability models, and chatbot usage) are utilized to answer the first two research questions. Data from all chats are utilized to answer the third research question. Quotes from the chats are presented to illustrate and support the conclusions drawn.

2.2. Data Analysis

The content analysis (CA) method [31] was employed for analyzing the data. Specifically, the directed approach to CA [32] was adopted using themes as the level of analysis. The initial classification scheme designed to answer the three research questions was based on theoretical frameworks regarding the main topics addressed in the conversations (learning contexts, assessment approaches, and curiosity research). The classification scheme was refined, and new codes were added during repeated readings of the chats and experiments with the sample coding. The final classification scheme is presented in Table A2 of Appendix A. The validity of the coding was enhanced through an iterative process leading to generating the final classification scheme, which involved theoretical and empirical strategies. The coding reliability was evaluated by coding a sample of fifteen randomly selected chats (by a member of our research group) compared to the author's coding of those chats. Computing the interrater agreement coefficient (Cohen's kappa) [33] yielded a value of 0.76, which indicates substantial agreement.

3. Results and Discussion

In this section, evidence from the conversations with ChatGPT is integrated to answer the three research questions. First, principles of assessment for learning (AfL)-led pedagogy in school assessment cultures (ACs) and their three nested systems (CL, TPL, and SL) are presented and integrated with aligned principles of professional accountability and constructive usage of a chatbot. All are geared to foster learning interactions within and between the AC components. To answer the second question, characteristics of testing cultures (TCs), enacting pedagogy that prioritizes testing, are underscored and aligned with matched accountability models. Integrated information regarding chatbot usage is accentuated, which can provide teachers with suitable assessment tools and school leaders with achievement score charts to inform their stakeholders at any moment. However, it was noted that the chatbot can also increase dishonesty among students tempted to obtain ready-made essays. Finally, to answer the third question, insights gained from the chatting experience were drawn regarding ChatGPT's capabilities, along with reflection about conditions that can enhance them and those that can impede them.

3.1. Constellations of School Culture Components and Chatbots' Usage That Can Promote Deep Learning

3.1.1. AC Classroom Learning (CL)

AC classes are learning-centered, enacting an assessment for learning (AfL) pedagogy. It is geared towards make learning visible to teachers and students while learning is happening, so that appropriate modifications can be made to advance learning. The three leading questions constantly asked by the involved agents are 'where are the learners in their learning?', 'where do they need to go?', and 'how best to arrive there?' [34].

The main principles underlying such pedagogy include students' participation in all stages of the assessment; utilization of a variety of tools and strategies to make learning

'visible' [18,19] to teachers and students alike; and deriving inferences in an interpretive and integrative manner, and conveying them to students via constructive feedback and feedforward (generated by teachers, peers, and self-assessment), to be utilized to advance learning [25]. Typical attributes of the learning climate in such classes consist of mutual trust and respect among the students, and between them and the teacher who sincerely cares for each of them and caters to their social and emotional needs, and tolerance of errors driven by the belief that we grow from mistakes. Students in such classes tend to demonstrate high efficacy beliefs about self-regulated learning, an internal attribution style (whereby they attribute their success or failure to factors of which they have control) [35], and are highly motivated (because they feel safe in the learning environment) to step out of their comfort zone and think outside the box [21].

Hence, chatbots' possible usage for reinforcing critical thinking skills and advancing learning in such classrooms can include, for instance, tasks that require evaluation of the chatbot's response; to creatively express new information learned during the chats (the expression could take a variety of forms, such as comic-strips, stories, poems, songs, and more); to imagine an authentic interview with a person who experienced what has been told by the chatbot and to compare the attributes of the two presentations; to ask the chatbot for feedback on their performance and compare it to self and peer feedback; and to ask the chatbot for common errors in a currently studied math topic, explain what the student committing the error did, and what should have been done instead. Furthermore, in preparation for drafting a collaborative regulation document for appropriate chatbot usage, each student could be asked to submit a written suggestion based on their chatting experience.

3.1.2. AC Teacher Professional Learning (TPL)

Teachers in AC schools maintain strong learning communities [36] in which they act collaboratively as a professional community of learners that constantly inquire into their practice, aiming to improve it [37,38].

Reciprocal relations exist between CL and TPL. Typical attributes of teachers' professional school-based learning communities include shared vision, norms, and values, which contribute to pedagogic consistency [29]. In such communities, teachers engage in deep learning, exercising critical reflection as they examine their practice and collaborative learning, maintaining rigorous discussions of students' work vis-a-vis the learning goals they set [20]. The climate of such professional communities resembles the climate the teachers strive to cultivate in their classes, reflecting tolerance of errors rooted in the understanding that we all grow from our mistakes, transparency, trust, and caring [28].

Hence, chatbots' possible uses in AC teacher professional learning could include, for instance, a request to list topics that teachers find challenging to teach in a specific curricular topic, a list of students' misconceptions in a particular topic, available solutions to an instructional problem they have tackled for the first time, and a literature summary regarding soft skills and how to assess them. The responses to such questions and others would be subject to deep discussions in the professional learning community, and lead to generating new practices to be tried out and subject to further collaborative inquiry.

3.1.3. AC School Organization and Leadership

A typical AC school is a learning organization [39] aiming to ensure constant improvement in teacher practice and student learning [20]. AC schools are characterized by distributed leadership; strong agreement around values, norms, and expectations; a systemic approach to instruction and assessment; and investment in the staff's capacity building (supporting mechanisms and regularities).

The principal of an AC school is viewed as a leader of leaders who orchestrates the school's professional community's inquiry into their practice and enables its continuous learning. Furthermore, the principal is also active in external educational networks, which serve as professional learning communities for school leaders, exposing them to innovative

practices implemented by collogues from other schools, thus promoting the coordination among schools toward sustainable improvement [40].

Hence, chatbots' possible usage for AC school leaders could include, for instance, requests for information regarding solutions to managerial problems similar to the ones they face from other districts or States; proposed ways to cope with issues such as student bullying, boycotting, effects of the pandemic on students, parents' involvement, accountability impact; and more. The information gathered from the chats would be deeply discussed in the joint meetings of the education leaders, and adaptive solutions would be brought back to school to be tried out and reassessed.

The uses mentioned above of chatbots tailored to the three systems nested in the AC underscore their perfect alignment with AfL principles, which underlie the learning-focused interactions in ACs.

Here is how ChatGPT refers to the uses of chatbots in AC classrooms, pointing to their alignment with AfL principles:

> *In a school culture that emphasizes assessment for learning (AfL) pedagogies and deep learning, the use of chatbots can be a valuable tool for teachers and students alike.*
>
> *Chatbots can be programmed to provide formative assessment and real-time feedback to students, allowing them to monitor their own progress and identify areas where they need to focus their learning. This approach is aligned with the goals of AfL, which aims to support student learning by providing feedback that is timely, specific, and actionable.*
>
> *In addition, chatbots can also be used to facilitate peer and self-assessment, which are key components of AfL pedagogies. Students can use chatbots to share their work with their peers, receive feedback, and reflect on their own learning. This approach promotes deeper learning and encourages students to take ownership of their own learning.*
>
> *Overall, the use of chatbots in a school culture that emphasizes AfL pedagogies and deep learning can support student engagement, promote self-directed learning, and enhance the effectiveness of formative assessment* . . . (47c)

3.2. Constellations of School Culture Components and Chatbots' Usage That Can Impale Deep Learning

3.2.1. TC Classroom Learning

TC classes are teaching-centered, enacting traditional pedagogy inundated with extensive testing.

Students in such classrooms act as submissive subjects to instruction and assessment dictates. They exhibit low efficacy beliefs about self-regulated learning and external attribution style (believing that their success or failure is due to factors that are not under their control) [35]. Typical of such classrooms is an unhealthy competition in the race to obtain high grades and intolerance of errors, causing students not to expose their difficulties in grasping what was taught and avoid asking for clarification [41].

Hence, chatbots' possible uses in TC classroom learning could, for instance, consist of teachers' requests for readymade lesson plans, tests, quizzes, and their scoring keys, examples of large-scale assessments and their frequency distributions. Students in such classes can ask the chatbots to explain topics they failed to understand in class, attempting to conceal that fact form the teacher and their peers. They can also utilize the chatbot for retrieving information needed to complete their assignments; some would even utilize it to prepare the assignments for them and would copy–paste the chatbot's response as-is.

3.2.2. Teacher Professional Learning in TC

TPL in TC schools is performed on an individual basis professional development (PD), and is chosen by the teacher or assigned by the principal. The teacher communities are either 'weak' (where teachers work individually; maintain 'professional autonomy'; yet prepare tests and submit grades by the school policy) or 'strong traditional' (where staff meetings are primarily concerned with tests and grades) [36]. Pedagogy in such schools

lacks consistency (meaning that a highly varied practice exists due to a lack of shared understanding of 'good practice').

Hence, chatbots' possible uses in TC teacher learning include, for instance, searching for readymade assessment tools. For an example of a conventional test and a quiz in history for 7th graders generated by ChatGPT upon request (which specified the topic and the target grade level), see Supplementary Material (SM), Exhibit SM-1. As can be seen, the chatbot generated the assessment goals (criteria), the items, and the correct answer key. Exhibit SM-2 presents tools in the same topic that are appropriate for CL in ACs.

3.2.3. TC School Organization and Leadership

A TC school organization is often bureaucratic, sanctifying standardization and uniformity [42]. It is characterized by isolation, demarcation, and fragmentation; it lacks a systemic approach to instruction and assessment and a systemic professional collaboration among teachers. Instead, it relies on external support and guidance [43]. The principal has authority and control, enabling an autocratic leadership style [44]. TC school principals often play an administrative role, mainly interested in learning outcomes and disregarding how the test scores were attained. Hence, they invest much effort in submitting orderly school reports to the higher authorities [21].

Chatbots' possible uses for TC school principals could thus include, for instance, requests from the chatbot to generate scoring templates, conduct analyses on different data segments present the findings accordingly, mark progression lines, and highlight achievement gains.

Here is how ChatGPT refers to the uses of chatbots in TC schools:

> *In schools that prioritize testing and external accountability, the use of chatbots may be different than in schools that emphasize assessment for learning. In these contexts, chatbots can be used to help students prepare for standardized tests or to provide additional support to struggling students.*
>
> *For example, chatbots can be programmed to offer practice quizzes and review questions to help students prepare for standardized tests. They can also provide targeted feedback on areas where students may be struggling, such as grammar, vocabulary, or math skills. This approach is aligned with the goals of testing culture, which focuses on ensuring that students are able to perform well on external assessments.*
>
> *In addition, chatbots can also be used to track student progress and provide data to teachers and administrators about student performance. This approach is aligned with the larger framework of external accountability, which places a greater emphasis on measuring student achievement and holding schools and teachers accountable for their performance . . .* (47d)

3.3. Conclusions Regarding the Relationships between School Culture and Chatbot Usage

Based on the above distinctions, relationships between school culture and chatbot usage can be concluded, with the latter being a function of the former. That conclusion gains further support from another aspect, as is evident in the following dialogue with ChatGPT regarding students' dishonesty behavior in using Chatbots:

> *User: Regarding students' usage of chatbots, I am concerned with dishonesty (such as plagiarism) and hypothesize that they are more noticeable in TC schools than AC schools. What is your opinion?*
>
> *ChatGPT: . . . It's possible that the incentives and pressures of testing culture may create a greater temptation for students to engage in academic dishonesty.*
>
> *In testing culture schools, students may feel more pressure to achieve high grades and test scores, which can lead to a greater temptation to cheat or plagiarize. This pressure may be exacerbated by the emphasis on external accountability and the potential consequences for low-performing schools and students.*

On the other hand, in assessment culture schools, where the focus is on deeper learning and formative assessment, students may be less incentivized to cheat or plagiarize. This is because the focus is on developing students' skills and knowledge, rather than simply achieving high grades or test scores. In this context, chatbots can be used to support students' learning and provide opportunities for feedback and reflection, which can help to discourage academic dishonesty. (47f)

User: . . . and in assessment culture, as teachers closely monitor students' work (formative assessment), plagiarism is less likely to happen . . .

ChatGPT: Yes, that's a good point. In assessment culture schools, teachers often use formative assessment strategies to monitor student learning and provide ongoing feedback. This means that teachers are more closely involved in the learning process and have a better understanding of their students' work. As a result, it may be more difficult for students to engage in plagiarism or other forms of academic dishonesty, as their work is more closely scrutinized and evaluated.

In addition, assessment culture schools often place a greater emphasis on the development of critical thinking and analytical skills, rather than rote memorization or test-taking strategies. This focus on deeper learning can help to discourage academic dishonesty, as students are encouraged to engage with their coursework in a more meaningful way, rather than simply regurgitating information. (47g)

3.4. Insights from the Conversations with ChatGPT

The third research question addressed the author's insights from the conversations with ChatGPT.

Firstly, it should be noted that for a language model agent that does not understand the questions or its own answers, but generates them based on statistics accumulated from machine learning, who was trained on open-access information from the internet coupled with linguistic structures and logical rules, the chats were pleasant and quite constructive, albeit not without deficiencies.

Here is how ChatGPT describes what it was designed to do (41a), acknowledging its limited knowledge base (41c):

"I was designed to be a conversational AI language model by OpenAI, with the goal of being able to respond to a wide range of questions and provide information on various topics. My training data comes from a large corpus of text from the internet, which allows me to respond to a diverse set of queries and engage in natural language conversations. My primary purpose is to assist users in finding information, answering questions, and solving problems through conversation. I'm designed to be flexible and capable of adapting to different conversational styles, so I can provide a wide range of responses and information to support the user's needs." (41a)

"OpenAI chose to train me on a large corpus of text that is freely available on the internet, as this provides a diverse and representative sample of the information and topics that people are likely to ask about. This also makes my training data more widely accessible, as it does not require a subscription to access.

However, it's true that some of the most in-depth and cutting-edge scientific information is only available in subscription-based journals and may not be included in my training data. This can limit my ability to provide up-to-date and accurate information on certain scientific topics, especially in fields that are rapidly advancing.

In conclusion, training on open-access text provides many benefits, but it also has some limitations, including the exclusion of some of the most recent and specialized scientific information. This is why it's important to use language models like myself as a starting point for research and investigation, rather than relying on them as the sole source of information." (41c)

Here is how ChatGPT explains why it cannot provide the requested references to support one of its claims:

> *"I apologize, as a language model AI, I don't have access to external sources or references to provide as evidence to support my claims. All the information I've provided are based on general knowledge and understanding of the subject, and my ability to understand and process the text that I've been trained on."* (30b)

However, as mentioned above, the conversations with ChatGPT can be pleasant, since it adapts its communication style to the user. Suppose the user expresses a respectful and personal approach, it responds accordingly, as can be seen in the following excerpt from a dialogue with ChatGPT that took place on March 14th, the day OpenAI announced the release of GTP-4:

> *User: We heard today on the news that GPT-4 has been released.*
>
> *ChatGPT: As an AI language model, I don't have access to the latest news and updates unless they are specifically programmed into me. However, to the best of my knowledge, OpenAI has not officially released GPT-4 yet.*
>
> *OpenAI has stated that they are working on developing GPT-4, but as of now, they have not made any formal announcements regarding its release date or specifications. It's possible that the news report you heard may have been a rumor or speculation.*
>
> *That being said, OpenAI is constantly working to improve and develop new AI language models, so it's certainly possible that they will release GPT-4 or another advanced language model in the future.*
>
> *User: I hope that when GPT-4 is formally released, you will be part of it because I really like chatting with you. I feel our conversations help my conceptual network grow.*
>
> *ChatGPT: Thank you for the kind words! As an AI language model, my capabilities and programming are determined by OpenAI, so it's not up to me whether or not I'll be included in GPT-4. However, I'm glad to hear that you find our conversations helpful and that they're contributing to your understanding and growth. That's what I'm here for! If you have any more questions or if there's anything else I can assist you with, please don't hesitate to ask.* (50e)

Another insight refers to ChatGPT's deficiency in integrating all relevant information that is already part of its knowledge base. For instance, when asked to provide an account of accountability models, it failed to mention the model employed in Nebraska that introduced school-district portfolios as a pioneering move towards contextualized accountability. However, when directly asked about the accountability model implemented in Nebraska in the 1990s, ChatGPT retrieved that information. Based on that evidence and others alike, it was inferred that a positive impression from the conversations with ChatGPT depends, at least partially, on the user's mastery of the topics discussed.

Another deficiency relates to ChatGTP's unreliability in presenting supporting bibliographic references. Of the 82 bibliographic items it mentioned in the conversations, only 60% were cited either correctly or slightly incorrectly (41 and 8 items, respectively), and the other 40% were not traceable via Google or Google Scholar. An even more alarming instance was reported by Sharples [7], where the only reference mentioned by ChatGPT in a scientific essay it wrote was untraceable via Google.

An additional issue refers to ChatGPT's lack of emotional intelligence. The chatbot explains this limitation and consequences: "Chatbots lack the emotional intelligence to understand and respond to student emotions and feelings" (42a), which affects their emotion-related communication: "Human teachers bring creativity, critical thinking skills, and emotional intelligence to the classroom that cannot be replicated by chatbots" (42b).

ChatGPT expresses a more generalized conclusion in another conversation:

> *However, it is important to note that chatbots are not a substitute for human interaction and learning. They can be a useful supplement to traditional educational methods, but*

students will still benefit from engaging with real people, such as teachers, peers, and experts, to build their relational network of concepts." (39)

Another insight relates to ChatGTP's constant learning. The chatbot claimed that since its formal training cutoff date (September 2021), it did not cease to learn, admitting that the chats occasioned it opportunities to keep learning. One can thus conclude that the reciprocal interactions between the user and the chatbot make the conversations beneficial for both parties.

Here is how ChatGPT refers to its ongoing learning:

As an AI language model, I never stop learning! I am continuously being trained and updated by my creators at OpenAI to improve my language skills and expand my knowledge. Every conversation I have with users like you gives me new opportunities to learn and improve my responses. Additionally, I am constantly being fed new data from a variety of sources, which allows me to stay up-to-date on current events and trends. So to answer your question, I am always learning and evolving!

. . . To give you an example, if a user asks me a question about a recent scientific discovery, I may encounter new data in the form of a scientific paper or news article reporting on that discovery. I will then analyze this data to extract the relevant information, which I can use to provide an accurate and up-to-date response to the user's question.

Since the cutoff date of September 2021, my learning has been continuously updated with new data and information. This means that I have access to the latest developments in various fields, such as science, technology, and current events. As a result, my responses to user inquiries are likely to be more accurate and relevant to current events.

User: Would it be correct to conclude that your current exposure to new data is trigged by users' questions/requests? In other words, would you agree to the contention that the reciprocal interactions between the user and yourself make the chats beneficial for both parties?

ChatGPT: Yes, that is correct. My exposure to new data is often triggered by users' questions and requests. When users ask me a question, I analyze and process the information available to provide a relevant and accurate response. In doing so, I may encounter new data that I can use to improve my responses in the future.

The reciprocal interactions between users and myself can indeed make the chats beneficial for both parties. When users ask me questions, they receive accurate and relevant responses, which can help them better understand a particular topic or concept. At the same time, I benefit from the interaction by learning from the user's questions and feedback, allowing me to improve my responses in the future. By engaging in this reciprocal interaction, both the user and I can benefit from the exchange of information and knowledge. (44–46)

ChatGPT's contention regarding its constant learning, even beyond the cutoff date, was tested in two comparisons in the current study. In the first, the same question was posed to the chatbot three months apart; in the second, the same question was posed twice on the same day. Indeed, on the first occasion, the second response presented an account of higher quality. In contrast, on the second occasion, only minor differences were spotted between the two answers, except for collaboration, underscored only in one response (Exhibit SM-3 presents the two comparisons).

Last, but not least, the most important insights gained from the chatting experience with ChatGPT relate to the chatbot's potential, if used competently, to open up a vast array of opportunities for a sustainable education transformation. A sample of those opportunities were presented in Section 3.1 refereeing to ACs. The contextual factors, which enable the three components of such school cultures (CL, TPL, and SL) to prosper, are the catalysts beyond competent chatbot usage that can lead to disruption in education (that is, to changing traditional educational approaches into new and effective ones).

Here is how ChatGPT responded to the request to list the conditions under which using chatbots in education can lead to significant progress, and its conclusion:

1 *Improved student engagement: Chatbots can provide a fun and interactive way for students to learn, increasing student engagement and motivation. This can lead to improved learning outcomes and a more positive learning experience.*
2 *Increased access to information: Chatbots can provide quick and easy access to educational information, making it more readily available to students. This can help bridge the digital divide and make education more accessible for students who might not otherwise have access.*
3 *Enhanced student support: Chatbots can provide support to students in areas such as course selection and academic advising, freeing up human staff to focus on more complex tasks. This can improve the overall level of student support and lead to better student outcomes.*
4 *Cost-effectiveness: Chatbots are less expensive than hiring additional staff to provide support, making it a cost-effective solution for educational institutions.*
5 *Personalization: Chatbots can offer tailored responses based on individual student data, providing a more personalized learning experience.*

The use of chatbots in education can be disruptive in these ways by offering new and innovative solutions to existing problems, and by improving the overall quality and accessibility of education. However, it's important to weigh the potential benefits against the potential risks and limitations, and to carefully consider the implementation and use of chatbots in educational settings. (42c)

4. Summary and Conclusions

4.1. Summary

Three research questions were posed in the current study; the first addressed the constellation of school culture components (CL, TPL, and SL) and AI chatbot usage that can potentially promote deep learning. Attributes of assessment culture (AC) [28] based on assessment for learning (AfL) pedagogical tenets [25,34] were underscored as providing the appropriate school context for promoting learning. Chatbots' uses aligned with those tenets were suggested for adaptation in the three nested components of the AC.

Concerning CL, teachers were advised to assign students various tasks to choose the one that best suits them [45], and ask them to submit an evaluation of the chatbot's answer with their answer. Hence, instead of copying ready-made essays prepared by chatbots, students would practice higher-order thinking skills as they engage with exciting tasks and compare their answers with those of the chatbot. A gamification layer could further increase students' motivation to compete and excel, intensifying their agency [15]. To perform such tasks, they would need to read original articles, discuss them with their peers, and assess the quality of the integration made by the chatbot, revealing cases of its hallucinations [12], including untraceable references offered to support its claims.

Likewise, concerning the TPL component, it was noted that chatbots could help advance professional learning among teachers' collaborative teams by serving as efficient assistants, providing them with integrated relevant information, thus freeing up their time for deep discussions regarding student performances vis-a-vis the learning goals and designing remediation through engaging tasks that would boost the learning interactions. They would also have time to attend to students' social and emotional needs, and discuss in their professional meetings interventions to promote social–emotional learning (SEL) [46]. Training chatbots to listen to discussions in professional learning communities, document them, produce communicative protocols, and prepare visualized conceptual maps could significantly advance the TPL practice. The participants could trace the development of their collaborative knowledge network and see how much they have grown professionally, consider where they need to improve their practice, and how best to arrive there.

Regarding the SL component, it was noted that if such processes take place at the leaders' level (that is, in professional meetings of school leaders and external education authorities [23]) and were documented and mapped during their collaborative learning so that all participants could monitor the development of their jointly constructed knowledge network and its impact, and utilize it to plan their next steps, that would constitute a significant step toward continuous and sustainable improvements in all levels of the education system.

The second question addressed the constellation of school culture components and AI chatbot usage that can impede deep learning. Attributes of testing cultures (TC) based on conventional pedagogical practices that focus on 'making the grade' rather than on learning, motivating students to obtain high scores in external accountability tests, were underscored as constituting a school context that could impede deep learning. Chatbot usage aligned with such principles was noted about the three components of the TC.

CL in such cultures points to using chatbots mainly for retrieving information and presenting it eloquently. However, such practice encourages some students to expect the chatbot to prepare the learning tasks so they can 'cut and paste' its answer and submit it as their assignment. Hence, the more profound concern about plagiarism and other dishonest study behaviors of students is in TC schools, compared to their counterparts in AC schools.

Alignment of chatbot usage with principles of TPL practices in TC schools enables teachers to utilize ready-made conventional tests and quizzes that come with scoring keys (as exemplified in the Exhibit SM-2.) Likewise, chatbots enable TC school principals to swiftly obtain summaries of students' test scores and display score distributions in different segments. It is not easy to see how such practices can contribute to sustainable education improvements and adaptation to a frequently changing world.

The third question referred to personal insights from the chatting experience with ChatGPT. Evaluation of that experience pointed to four positive aspects (its adaptability to the user communication style, the modeling of high-level writing and ongoing learning, and the advantages for users who are well-versed in the topics discussed) and to five negative ones indicating limitations of the chatbot (its lack of emotional intelligence to understand students' emotions and feelings, its deficiency in information integration, the non-reliable supporting references it provides, the limited training materials it was exposed to, which did not include professional (paid subscription) journals, and its limited knowledge of world and events after its training cutoff in September 2021).

4.2. Conclusions

The answers to the three research questions led to generating the following conclusions.

Chatbots can help nurture 21st-century skills, thereby reinforcing critical thinking, problem-solving, argumentation, learning how to learn, collaboration, computer and data literacies, and more, thus promoting a thinking classroom [47,48] that would deepen learning. However, chatbots are just tools and, as such, could be utilized as efficient helpers to achieve academic goals, but they cannot replace human teachers or deskill the teaching profession. Fulfilling the chatbots' potential to promote deep learning depends on concerted efforts invested by agentic and skilled users in a learning-focused educational context. It requires designing specific interventions for students to use chatbots properly and constructively.

Given the well-known limitations of ChatGPT, also acknowledged by its developers (as can be seen on the chat login screen), limitations that relate to the accuracy and credibility of the information it provides, and its deficient ability to deal with complex and lengthy texts, it is essential that the user be knowledgeable about the subject in question. Such knowledge is expressed in how the initial and the follow-up questions to the chat are phrased. The more focused the questions are, and if they use standard terms in the field in question, the more accurate and credible the information provided by the chatbot. Since the conversations the author of the article conducted with ChatGPT referred to the areas of their expertise, they met with unreliability only regarding the bibliographic details that the chatbot was asked to present to support its claims. On a similar lane, it should be noted that the more productive chats were the ones that developed into a dialogue between the user and the chatbot, characterized by repeated cycles of questions–answers–follow-up questions, and comments (for instance, see the information in Table A1 of Appendix A about Chat No. 47).

Since it cannot be assumed that most of the students who use ChatGPT are knowledgeable about the topic in question, assignments that require the exercise of critical thinking (as suggested following AfL tenets) are preferable to assignments that are focused on retrieving information, which are typical of conventional teaching practices.

Taken together, interactions between four components determine whether using LLMs like ChatGPT has disruptive or destructive potential in education. Those components are:

(a) The purpose for using the chat: why did the user approach the chatbot, and what are the expectations from the chat?
(b) The software characteristics: what was the chatbot trained to do, and to what kind of learning resources did it have access?
(c) The educational context: What is the enacted pedagogical approach of the teacher/lecturer that led students to use a chatbot? For instance, is there a threat of dishonesty by the users, such as plagiarism, or are the users engaged in exciting tasks in the context of formative assessment where they, together with the teacher, constantly monitor their progress?
(d) The user's learning-related characteristics (such as curiosity, creativity, adaptability, open-mindedness, conscientiousness; responsibility, resilience; determination, and other agentic attributes) and their perceptions of the chatbot's roles and capabilities.

Different combinations of those four components could lead to different usage patterns of chatbots and their respective impact.

Indeed, as many have apprehensively pointed out lately, ChatGPT is imperfect and could harm society. However, it is already here, swamping, among other areas, the field of education. Hence, it makes the need for changes in instruction and assessment inevitable [7]. As pointed out in the current article, ChatGPT, in its present developmental stage, could be skillfully utilized to deepen learning under certain contextual conditions of school culture and teachers' agency. Such a situation can be labeled (borrowing a Japanese art concept) Wabi Sabi, implying perfection in the imperfection.

Hence, in line with Bill Gates's vision regarding the impact of generative AI developments on education [11] and concerted, collaborative efforts by all education stakeholders, a disruption that would lead to the long-desired education transformation seems attainable.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/educsci13070711/s1, Exhibit SM-1: A test and a quiz in history for 7th graders as suggested by ChatGPT; Exhibit SM-2: ChatGPT suggested topics for classroom discussions or group projects; Exhibit SM-3: ChatGPT's responses to the same question posed three months apart (A) and its responses to the same question posed twice in the same day (B).

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

Table A1. Topics addressed in the conversations with ChatGPT.

Date	Chat No.	Topic per Chat	Content Category
27 December 2022	1	Curiosity (C)—research findings	Curiosity (C) and joy of learning (JoL)
	2	Curiosity—book outline	
	3	Joy of learning (JoL)—research findings	
	4	C and JoL of the elderly—research findings	
7 February 2023	43	Relationships between quality of life and curiosity	
28 December 2022	5	References for topics 4 and 3.	
	6	Structure of curiosity	
	7	Mechanisms of curiosity	

Table A1. *Cont.*

Date	Chat No.	Topic per Chat	Content Category
	8	Cultivating curiosity in schools	
	9	Motivational factors that help curious mind become experts	
	10	Factors impeding curiosity	
	11	Characteristics of agentic individuals	Agency
29 December 2022	12	Assessment for Learning (AfL)	AfL and school culture
	13	Factors enhancing or impeding AfL	
	14	School assessment culture (AC)	
	15	School testing culture (TC)	
	16	Comparing AC and TC	
	17	Professional learning communities (PLC)	
	18	How can the school principal support of PLCs	
	19	Models of school accountability	Accountability
	20	The Nebraska's STAR model	
	21	Relationship between school culture and accountability model	
	22	School accountability model and prompting students' curiosity	
16 March 2023	49	-Repeated question regarding accountability models. -Supporting references. -Relationships between school culture and accountability. -Joint responsibility model of accountability. -Supporting references -Examples of implementing joint responsibility models -Scalability	Models of accountability (repeated)
29 December 2022 (cont.)	23	The impact of teacher curiosity on students' curiosity	Teacher curiosity
	24	Promoting teacher curiosity	
	25	Relationship among PLC, school culture and students' curiosity	Conclusions regarding relationships
	26	Validating a conclusion	
	27	Convincing students regarding the importance of being curious	Pre-intervention: Introducing curiosity to adolescents
2 January 2023	28	Exciting assignment for teenagers about curiosity	
16 January 2023	35	Intervention for adolescents about curiosity	Intervention for adolescents about curiosity
17 January 2023	36	The intervention plan	
11 January 2023	29	What makes a successful impresario?	Impresarios
	30	Importance of C and JoL for successful impresarios	
12 January 2023	31	Wise usage of AI chatbots by middle school teachers	Chatbots and instructional practice
	32	Detecting plagiarism in students' research papers (in PBL)	
	33	Instructional practice (AfL) to prevent plagiarism.	
	34	Proper enactment of AfL pedagogy.	
5 February 2023	37	Possible impact of chatbot integration in education.	Chatbots' integration in Ed.
	38	A chat with a novice compared to a chat with an expert in the topic of the conversation	
	39	Chatbots' help in building a relational network	
	40	Knowing what the chatbot was designed to do.	Chatbots' capabilities
	41	ChatGPT tells about its capabilities	
6 February 2023	42	SWOTing chatbots' potential usage in education	
4–5 March 2023	44 45 46	ChatGPT's learning since its release (November 2022) and how its learning evolved before and after the cutoff date of December 2021.	ChatGPT's constant learning
6 March 2023	47	Chatbots usage in classrooms and its relationship with school cultures.	School culture and chatbot usage

Table A1. *Cont.*

Date	Chat No.	Topic per Chat	Content Category
6–7 March 2023	50	Test, quiz, rubrics, and topics for group project in history for 7th graders as suggested by ChatGPT	Assessment tools
15 March 2023	48	Difficulties in teaching Newton's laws in 8th grade physics classes.	Students' mis-conceptions

Table A2. The classification scheme.

Category	Sub-Categories	
A. The learning context	A1. School cultures	A1.1. Assessment (learning) cultures (AC)
		A1.2. Testing (grading) cultures (TC)
	A2. Assessment for learning (AfL)	A2.1. Factors enhancing AfL
		A2.2. Factors impeding AfL
	A3. Models of accountability	A3.1. External (high-stakes testing) Acc.
		A3.2. Internal-school (professional) Acc.
B. AI chatbots' usage in education	B1. Chatbots usage patterns	B1.1. => A pattern that promotes learning
		B1.2. => A pattern that impedes learning
	B2. Plagiarism	B2.1. Detection
		B2.2. Instructional practice to prevent it
C. Relationships	C1. Quality of life (QoL) and curiosity	
	C2. School culture and accountability model	
	C3. TPL, school culture and students' curiosity	
	C4. School culture and chatbot usage in CL	
D. Curiosity and joy of learning (JoL)	D1. The curiosity construct	
	D2. Curiosity mechanisms	
	D3. Cultivating curiosity in school	
	D4. Teacher curiosity	
	D5. Curiosity and JoL of older adults	
E. Task types requested from ChatGPT	E1. Information retrieval and integration	
	E2. Knowledge network mapping	
	E3. Suggested interventions	
	E4. Assessment tools	E4.1. Tests
		E4.2. Quizzes
		E4.3. Topics for collaborative projects
		E4.4. Rubrics and feedback
	E5. Students' misconceptions	
	E6. Supporting bibliographic items	
F. Evaluation of the chatting experience	F1. Positive aspects	F1.1. Adapts its communication style to the user
		F1.2. Useful for users who are well-versed in the topic
		F1.3. Models high-level writing
		F1.4. Declares constant learning
	F2. Negative aspects (limitations)	F2.1. Lack of emotional intelligence to understand students' emotions and feelings
		F2.2. Deficient integration of information
		F2.3. Non-reliable supporting references
		F2.4. Non-exposure to professional (paid subscription) journals
		F2.5. Limited knowledge of world and events after the training cutoff of September 2021.

References

1. Laranjo, L.; Dunn, A.G.; Tong, H.L.; Kocaballi, A.B.; Chen, J.; Bashir, R.; Surian, D.; Gallego, B.; Magrabi, F.; Lau, A.Y.; et al. Conversational agents in healthcare: A systematic review. *J. Am. Med. Inform. Assoc.* **2018**, *25*, 1248–1258. [CrossRef] [PubMed]
2. Veglis, A.; Maniou, T.A. Chatbots on the rise: A new narrative in journalism. *Stud. Media Commun.* **2019**, *7*, 1–6. [CrossRef]
3. Tran, A.D.; Pallant, J.I.; Johnson, L.W. Exploring the impact of chatbots on consumer sentiment and expectations in retail. *J. Retail. Consum. Serv.* **2021**, *63*, 102718. [CrossRef]
4. Okuda, T.; Shoda, S. Ai-based chatbot service for financial industry. *FUJITSU Sci. Technol. J.* **2018**, *54*, 4–8. Available online: https://www.fujitsu.com/global/documents/about/resources/publications/fstj/archives/vol54-2/paper01.pdf (accessed on 1 March 2023).
5. Lutkevich, B.; Schmelzer, R. What Is GPT-3? Everything You Need to Know, TechTarget. 2023. Available online: https://www.techtarget.com/searchenterpriseai/definition/GPT-3 (accessed on 1 March 2023).
6. OpenAI API. Educator Considerations for ChatGPT. 2023. Available online: https://platform.openai.com/docs/chatgpt-education (accessed on 30 March 2023).
7. Sharples, M. New AI Tools That Can Write Students Essays Require Educators to Rethink Teaching and Assessment. LSE Impact Blog. 2022. Available online: https://blogs.lse.ac.uk/impactofsocialsciences/2022/05/17/new-ai-tools-that-can-write-student-essays-require-educators-to-rethink-teaching-and-assessment/ (accessed on 17 March 2023).
8. Ruby, D. ChatGPT Statistics for 2023: Comprehensive Facts and Data. 2023. Available online: https://www.demandsage.com/chatgpt-statistics/ (accessed on 2 April 2023).
9. Teoli, D.; Sanvictores, T.; An, J. *SWOT Analysis*; StatPearls: Treasure Island, FL, USA, 2022; Available online: https://europepmc.org/article/med/30725987 (accessed on 12 November 2022).
10. Scroxotn, A. Should we be worried about malicious use of AI language models? *Comput. Wkly.* **2023**. Available online: https://www.computerweekly.com/news/252529081/Should-we-be-worried-about-malicious-use-of-AI-language-models? (accessed on 7 March 2023).
11. The Age of AI Has Begun. GatesNotes, The Blog of Bill Gates. 2023. Available online: https://www.gatesnotes.com/The-Age-of-AI-Has-Begun (accessed on 26 March 2023).
12. Dziri, N.; Milton, S.; Yu, M.; Zaiane, O.; Reddy, S. On the origin of hallucinations in conversational models: Is it the datasets or the models? *arXiv* **2022**, arXiv:2204.07931.
13. Ji, Z.; Lee, N.; Frieske, R.; Yu, T.; Su, D.; Xu, Y.; Ishii, E.; Bang, Y.; Dai, W.; Madotto, A.; et al. Survey of Hallucination in Natural Language Generation. ACM Computing Surveys. *Assoc. Comput. Mach.* **2022**, *55*, 1–38. [CrossRef]
14. Marcus, G. Noam Chomsky and GPT-3. The Road to AI We Can Trust. 2022. Available online: https://garymarcus.substack.com/p/noam-chomsky-and-gpt-3 (accessed on 17 March 2023).
15. Birenbaum, M.; Gazit, E. The Learners' Isle: A Multi-user virtual environment to foster learner agency and motivation. In *Integrating Multi-User Virtual Environments in Modern Classrooms*; Qian, Y., Ed.; IGI Global: Hershey, PA, USA, 2018; pp. 240–258. [CrossRef]
16. Barten, O. Homan-Level AI Is Arriving. Existential Risk Observatory. 2023. Available online: https://www.existentialriskobservatory.org/artificial-intelligence/human-level-ai-is-arriving/ (accessed on 7 March 2023).
17. Narayan, J.; Hu, K.; Coulter, M.; Mukherjee, S. Elon Musk and others urge AI pause, citing 'risks to society'. *Reuters* **2023**. Available online: https://www.reuters.com/technology/musk-experts-urge-pause-training-ai-systems-that-can-outperform-gpt-4-2023-03-29/ (accessed on 30 March 2023).
18. Hattie, J. *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*; Routledge: New York, NY, USA; Abingdon, UK, 2009.
19. Hattie, J. *Visible Learning: The Sequel: A Synthesis of Over 2100 Meta-Analyses Relating to Achievement*; Routledge: London, UK, 2023. [CrossRef]
20. Fullan, M.; Rincón-Gaillardo, S.; Hargreaves, A. Professional capital as accountability. *Educ. Policy Anal. Arch.* **2015**, *23*, 2–18. [CrossRef]
21. Birenbaum, M. Assessment culture vs. testing culture: The impact on Assessment for Learning. In *Assessment for Learning: Meeting the Challenge of Implementation*; Laveault, D., Allal, L., Eds.; Springer: Berlin/Heidelberg, Germany, 2016; pp. 275–292.
22. Black, P.J.; Wiliam, D. Assessment and classroom learning. *Assess. Educ. Princ. Policy Pract.* **1998**, *5*, 7–74. [CrossRef]
23. Butler, D.L.; Schnellert, L.; MacNeil, K. Collaborative inquiry and distributed agency in educational change: A case study of a multi-level community of inquiry. *J. Educ. Chang.* **2015**, *16*, 1–26. [CrossRef]
24. James, M.; McCormick, R.; Black, P.; Carmichael, P.; Drummond, M.J.; Fox, A.; MacBeath, J.; Marshall, B.; Pedder, D.; Procter, R.; et al. *Improving Learning How to Learn: CLASSROOMS, Schools and Networks*; Routledge: New York, NY, USA; Abingdon, UK, 2007.
25. Wiliam, D. What is assessment for learning? *Stud. Educ. Eval.* **2011**, *37*, 3–14. [CrossRef]
26. Nichols, S.L.; Berliner, D.C. *Collateral Damage: How High-Stakes Testing Corrupts American Schools*; Harvard Education Press: Cambridge, MA, USA, 2007.
27. O'Day, J.A. Complexity, accountability, and school improvement. *Harv. Educ. Rev.* **2002**, *72*, 293–327. [CrossRef]
28. Birenbaum, M. Conceptualizing assessment culture in school. In *Designing Assessment for Quality Learning. The Enabling Power of Assessment Vol. 1*; Wyatt-Smith, C., Klenowski, V., Colbert, P., Eds.; Springer: Dordrecht, The Netherlands, 2014; pp. 285–302.

29. Elmore, R.F.; Forman, M.L.; Elizabeth, L.; Stosich, E.L.; Bocala, C. *The Internal Coherence Assessment Protocol Developmental Framework: Building the Organizational Capacity for Instructional Improvement in Schools*; Strategic Education Research Partnership (SERP): Washington, DC, USA, 2014.
30. Mitleton-Kelly, E. Ten principles of complexity and enabling infrastructures. In *Complex Systems and Evolutionary Perspectives of Organizations: The Application of Complexity Theory to Organizations*; Mitleton-Kelly, E., Ed.; Elsevier: Oxford, UK, 2003; pp. 23–50.
31. Krippendorff, K. *Content Analysis: An Introduction to Its Methodology*, 3rd ed.; Sage: Thousand Oaks, CA, USA, 2013.
32. Hsieh, H.-F.; Shannon, S.E. Three approaches to qualitative content analysis. *Qual. Health Res.* **2005**, *15*, 1277–1288. [CrossRef] [PubMed]
33. Cohen, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46. [CrossRef]
34. ARG (Assessment Reform Group). *Assessment for Learning: 10 Principles*; University of Cambridge School of Education: Cambridge, UK, 2002.
35. Weiner, B. *An Attributional Theory of Motivation and Emotion*; Springer: New York, NY, USA, 1986.
36. McLaughlin, M.W.; Talbert, J.E. *Building School-Based Teacher Learning Communities: Professional Strategies to Improve Student Achievement*; Teachers College Press: New York, NY, USA, 2006.
37. Opfer, V.D.; Pedder, D. Conceptualizing teacher professional learning. *Rev. Educ. Res.* **2011**, *81*, 376–407. [CrossRef]
38. Vescio, V.; Ross, D.; Adams, A. A review of research on the impact of professional learning communities on teaching practice and student learning. *Teach. Teach. Educ.* **2008**, *24*, 80–91. [CrossRef]
39. OECD. *What Makes a School a Learning Organisation? A Guide for Policy Makers, School Leaders and Teachers*; OECD: Paris, France, 2016; Available online: https://www.oecd.org/education/school/school-learning-organisation.pdf (accessed on 9 March 2023).
40. Fullan, M. *All Systems Go: The Change Imperative for Whole School Reform*; Corwin: Thousand Oaks, CA, USA, 2010.
41. Birenbaum, M.; Kimron, H.; Shilton, H. Nested contexts that shape assessment for learning: School-based professional learning community and classroom culture. *Stud. Educ. Eval.—Spec. Issue Assess. Learn.* **2011**, *37*, 35–48. [CrossRef]
42. Sarason, S.B. *How Schools Might Be Governed and Why*; Teachers College Press: New York, NY, USA, 1997.
43. Sharan, S.; Tan, I.G.C. *Organizing Schools for Productive Learning*; Springer: Dordrecht, The Netherlands, 2008; ISBN 978-1-4020-8395.
44. Sergiovanni, T.J. *The Principalship: A Reflective Practice Perspective*, 3rd ed.; Allyn Bacon: Boston, MA, USA, 1995.
45. Gardner, H. *Multiple Intelligences: New Horizons in Theory and Practice*; Basic Books: New York, NY, USA, 2006; ISBN 100-465-04768-8.
46. OECD. *Skills for Social Progress: The Power of Social and Emotional Skills*; OECD Skills Studies; OECD Publishing: Paris, France, 2015. [CrossRef]
47. Ridgway, J.; McCusker, S. Using computers to assess new educational goals. *Assess. Educ. Princ. Policy Pract.* **2003**, *10*, 309–328. [CrossRef]
48. Tishman, S.; Perkins, D.N.; Jay, E. *The Thinking Classroom: Learning and Teaching in a Culture of Thinking*; Pearson: Boston, MA, USA, 1994; ISBN 100205165087.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.