International Electronic Journal of Mathematics Education
2023, 18(3), em0739
e-ISSN: 1306-3030
https://www.iejme.com

MODESTUM

Review Article

OPEN ACCESS

Designing mathematics standards in agreement with science

JudithAnn R. Hartman [1], Sarah Hart [2], Eric Alan Nelson [3]*, Paul A. Kirschner [4,5]

[1] Department of Chemistry (Emerita), United States Naval Academy, Annapolis, MD, USA
[2] Broad View Elementary School, Oak Harbor, WA, USA
[3] Fairfax County Public Schools (Retired), Fairfax, VA, USA
[4] Faculty of Educational Sciences (Emeritus), Open University of the Netherlands, Heerlen, THE NETHERLANDS
[5] Expertise Center for Education and Learning, Thomas More University of Applied Sciences, Antwerp, BELGIUM
***Corresponding Author:** EANelson@chemreview.net

Citation: Hartman, J. R., Hart, S., Nelson, E. A., & Kirschner, P. A. (2023). Designing mathematics standards in agreement with science. *International Electronic Journal of Mathematics Education, 18*(3), em0739. https://doi.org/10.29333/iejme/13179

ARTICLE INFO

Received: 03 Feb. 2023

Accepted: 05 Apr. 2023

ABSTRACT

To learn mathematics, historically students had no choice but to memorize fundamental facts and apply memorized algorithms. Since 1995 in the US, all states have adopted standards to govern K-12 mathematics instruction, and in most, standards have de-emphasized memorization and emphasized reasoning based on concepts. This change assumed the brain could reason in mathematics without relying on memorized knowledge. Scientists who study the brain have recently verified this assumption was mistaken. Due to stringent limitations in working memory (where the brain solves problems), mathematical problem-solving of any complexity requires applying well-memorized facts and procedures. A decade after the implementation of standards in most states, US young adults ranked last in testing in mathematics among 22 nations. Changes are proposed to state K-12 standards, which recent scientific research suggests would substantially improve student mathematics achievement.

Keywords: cognitive science, K-12 instruction, US K-12 mathematics standards, working memory limits, grade 1-4 instruction

INTRODUCTION

Mathematics is the foundation for the skills of a nation’s workforce in many areas, but especially in the sciences, engineering, and technology. For the mathematics skills of US students, test scores present a grim picture.

OECD Testing

On a 2012 international test of numeracy skills by Organization for Economic Co-operation and Development (OECD) for citizens aged 16-24 and 16-34, the US ranked last among 22 developed nations (Goodman et al., 2015; OECD, 2012) (**Figure 1**).

NAEP LTT

Since the 1970s, the National Assessment of Educational Progress Long-Term Trend test (NAEP LTT) has been given at multi-year intervals to large samples of students at ages 9 and 13 in all US states. The most recent LTT administrations were in 2012 and January 2020 (prior to US COVID-19). For age 9 students, comparing 2012 to 2020, math scores declined slightly. For 13-year-olds, overall scores declined by a statistically significant amount (**Figure 2**). Except for the top 10% in 2012 and 2020, age 13 scores declined. The largest declines were seen with lower achieving students and black and Hispanic students (National Center for Education Statistics [NCES], 2021). A news headline (Mahnken, 2021) summarized:

> “Long-term NAEP scores for 13-year-olds drop for first time since testing began in 1970s–’A matter for national concern,’ experts say.”

State Standards

These test results followed a period of changes in mathematics teaching in conjunction with the adoption of K-12 state standards. In 1989, state governors endorsed ‘K-12 standards for instruction’ to be adopted by each state, which local school districts would be required to follow. The same year, NCTM, an organization of education school professors, K-12 curriculum specialists, and teachers of K-12 mathematics, issued recommendations on the content of state standards.

Copyright © 2023 by Author/s and Licensed by Modestum. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

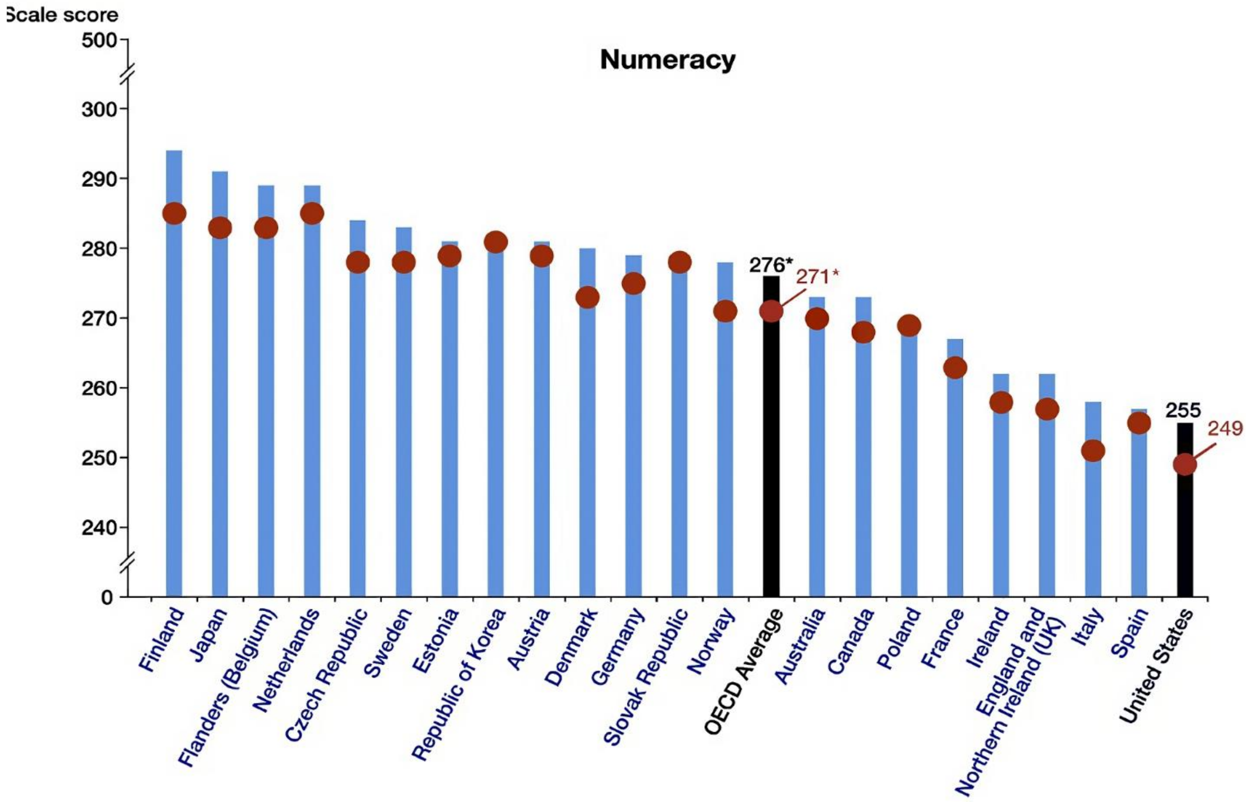

Figure 1. Average scores in numeracy for age 16-34 (millennials) & age 16-24 by participating country/region: 2012 (Goodman et al., 2015, p. 12)

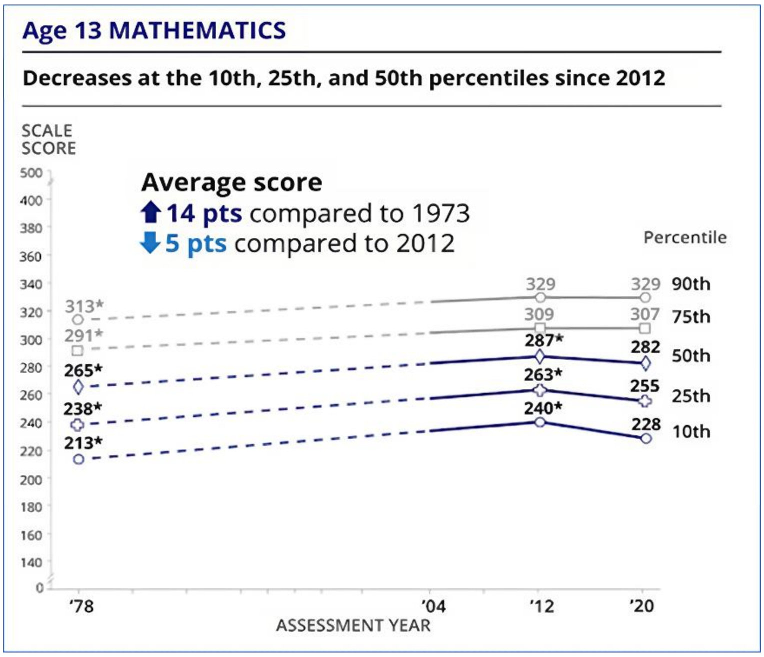

Figure 2. US scores by percentile, 1978 to 2020, NAEP LTT (Mahnken, 2021; NCES, 2021).

The NCTM's 1989 standards advocated calculator use "at all grade levels" and "increased attention" to "reasoning" to solve problems. "Decreased attention" was recommended for "memorizing rules and algorithms" and "rote practice" (NCTM, 1989, p. 20-21, 70-71). The NCTM standards made an unstated and untested assumption that when solving problems, the student brain could apply non-memorized, looked up, and mentally or 'calculator calculated' facts as easily as it applied facts that prior to the 1980's students had been required to memorize.

These changes impacted instruction. Between 1992 and 1997, over 45 states adopted and gradually implemented K-12 standards (Raimi & Braden, 1998). The federal No Child Left Behind Act of 2001 required all states to both adopt math standards and test on them annually in grades 3-8. A 2003 report by Loveless (2003) found "more than 40 states model their state math standards on those of the NCTM" (p. 175). By 2005, over 30 states called for calculator use in 3rd grade state testing (Klein et al., 2005).

Over 40 states in 2010-11 adopted and in 2014 implemented the 'Common Core math standards' (CCMS). Since 2014, a number of states have revised parts of their standards, but in most, standards remain quite similar to the 2010 CCMS (Norton et al., 2017). By the time of the 2020 NAEP LTT testing, the Common Core and similar standards had been required to be used in the classrooms of most states for more than five years.

Cognitive Science

During the period standards were being adopted, with assistance from new computerized technologies, scientists who study the brain were conducting research on how the brain solves mathematics problems.

Cognitive science is the study of how the human brain thinks, learns, and solves problems. Disciplines contributing to cognitive science include but are not limited to neuroscience, evolutionary biology, and cognitive and educational psychology. Prior to 1995, what could be said with scientific certainty about how the brain works was limited, but since then, with help from new technologies, scientists have been able to measure many of the brain's observable characteristics (Delazer et al., 2004; Dhond et al., 2003). In 2017, neuroscientist Seidenberg (2017) summarized, "although we are far from understanding how the brain works, many of its fundamental properties are known" (p. 138).

This article will review findings of cognitive research that address how the brain learns mathematics and how this relates to current teaching practices and mathematics standards. Except where noted, citations are either from research by cognitive scientists or from summaries of research written for educators by cognitive experts that limit technical terminology but include extensive references to peer-reviewed studies. A sample of current math standards in US states are also examined for their alignment with best practices identified by cognitive research.

TYPES OF LEARNING

We, as individuals and as a species, must solve problems to survive. As a result of natural selection, programmed into our DNA is the ability to automatically solve problems such as breathing and walking upright. In addition to these universal abilities, to enable us to adapt to a wide variety of environments, customized knowledge and skills based on our experience can be *learned*; that is, stored in our *long-term memory (LTM)* (Geary, 2002; Dehaene, 2020, p. 25).

Primary Learning

Over thousands of generations, human learning in some topics has evolved to occur without conscious effort. During 'window periods of development' (also known as 'sensitive periods'), for *limited* topics, the brain instinctively and automatically stores in LTM new knowledge gained from experience (Geary, 2002). As one example, creating speech is a task of enormous cognitive complexity, yet simply by exposure, virtually all human children become fluent in speaking the language they hear spoken around them (Pinker, 2007).

For additional limited topics including facial recognition, conventions of familial and social relationships (evolutionary psychologists term this 'folk psychology'), and a practical 'folk physics' and 'folk biology' of how things work, the LTM of children almost effortlessly fills with extensive knowledge. Learning that occurs automatically is termed *primary learning* (Geary & Berch, 2016).

Secondary Learning

To focus on survival, the brain evolved to put less focus on other learning. As one example, skill in reading, writing, and mathematics (also known as *cultural artifacts*) played no role in the long-term survival of our species. As a result, in those topics, learning is not automatic: It requires instruction, effort at recall (effortful learning), and practice. Such learning is termed *secondary*. The central purpose of schools is to structure secondary learning: to teach students how to solve problems arising in modern societies they do not learn to solve automatically (Geary, 2002; Sweller, 2008).

For a primary topic, after the window period closes, learning requires significantly more effort (Dehaene, 2020, p. 103-108). As one example, after about age 17, most of us lose the ability to learn to automatically apply the grammatical rules of a second language based on exposure alone (Hartshorne et al., 2018).

In disciplines such as history, economics, and law, domain knowledge is a mix of facts (such as names and dates) that are *secondary* (require effort to learn) with knowledge of human behavior that is *primary* (learned automatically as 'folk psychology'). In mathematics, however, minimal knowledge enters LTM without effort to make it recallable (Sweller, 2008).

PROBLEM CATEGORIES

Among scientists who study secondary learning, a 'problem' is broadly defined as "cognitive work that involves a moderate challenge" (Willingham, 2009a, p. 9). A math problem can range from "what does a + sign mean?" to complex calculus. Knowledge, problems, and problem-solvers separate into categories.

Verbatim vs. Gist

V*erbatim* (precise) and *gist* (summary) learning differ, and the brain treats them differently. Gist learning stores the basic concept of a larger quantity of knowledge. As an example, when listening to speech, to manage comprehension, the brain tends

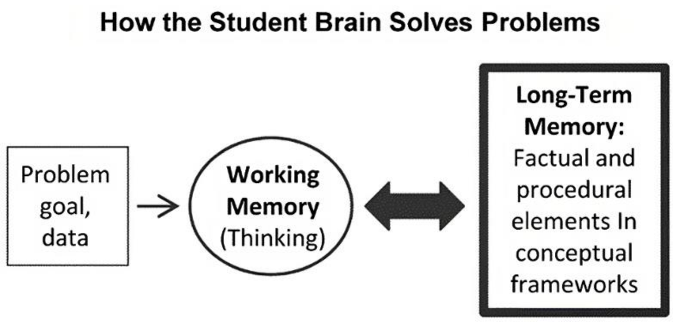

Figure 3. The (simplified) steps of information processing (adapted from Atkinson & Shiffrin, 1968)

to store a summary rather than "word for word recall" of what it hears. In contrast, in mathematics most facts that must be stored are verbatim (6×7 is not approximately 40, but rather 42). In the brain, verbatim facts are "encoded separately" from gist memories, and a precise fact "often requires more effort to learn than the gist … Verbatim recall … requires a great deal of time, effort, and practice" (Geary et al., 2008, p. 4-9).

Well- vs. Ill-Structured

Mathematics problems can be *well-structured* or *ill-structured*. Well-structured problems have a specific goal, initial state, constraints, and precise answers that can be found by applying structured procedures (*algorithms*)**.** All other types of problems can be categorized as ill-structured, including those for which correct answers are debatable or not known (Jonassen, 1997; King & Kitchener, 1994). In pre-graduate-level math courses, problems assigned to students are nearly always well-structured.

Expert vs. Student

Problem *solvers* can be divided into *experts* (including teachers) and *novices* (students). After years of study of a secondary topic, the LTM of experts contains a vast storehouse of well-organized knowledge. This enables experts to solve problems in their field by generalized reasoning strategies in ways students cannot (de Groot, 1946/2014; Kirschner & Hendrick, 2020, p. 4-12; Willingham, 2009c, p. 98).

In primary topics including social norms, folk biology and folk physics, adults are able to solve some problems using generalized reasoning (often some form of trial-and-error or weak problem-solving approaches) because our LTM has filled automatically with some topic knowledge (Anderson, 1987). But for complex problems in secondary topics including mathematics, non-experts (students) cannot solve by reasoning in the way experts can (Sweller et al., 2010).

For the remainder of this article, except where noted, we limit the scope to how the brain of students learns to use mathematics as a tool to solve well-structured problems.

WORKING MEMORY LIMITS AND STRENGTHS

Cognitive science explains how the brain solves problems based primarily on the interaction of three of the brain's components: the *attention* filter, *working memory (WM)*, and long-term memory (LTM) (**Figure 3**).

Storing Information

In working and long-term memory, elements of information are stored in networks of specialized cells called *neurons*. Each neuron can connect to other neurons and share information (Furst, 2020). Information is *encoded* in neurons in ways that allow it to be *remembered*: retrieved from LTM into WM and applied to solve problems (**Figure 3**) (Dehaene, 2020, p. 89-92; Trafton, 2017).

To share stored information, a neuron *fires*, creating an electrical signal that can travel to connected neurons. Thousands of fibers carry incoming signals to and outgoing signals from the nucleus of each neuron. Synapses are narrow gaps between the fibers from two neurons. When a fired outgoing signal reaches a synapse, molecules termed neurotransmitters are released, cross the gap, and may cause the adjacent neuron to fire. Together, these nuclei, fibers, and synapses are termed the brain's *wiring* (Dehaene, 2020, p. 10).

WM Overview

WM has multiple functions. It is a temporary memory that can focus on the limited information we *attend to* and are *conscious of* at the moment, such as unique problem data. WM can be described as composed of *slots* that can briefly store *chunks* of encoded information (Luck & Vogel, 2013; Willingham, 2006).

WM is where the brain solves problems. To solve a math problem, WM can use its stored problem-data to search LTM for matching data. By a process termed *cued recall*, additional data connected to located matching data by wiring can be *retrieved* from LTM into WM. The retrieved relationships can then be used to *process* problem data, converting the data step-by-step to reach the problem goal (Gathercole & Alloway, 2004). During problem solving, processing in WM can be observed as "the vigorous firing of many neurons" primarily in the brain's cortex (Dehaene, 2020, p. 90, 150).

WM is also where learning begins: New and known information processed at close to the same time in WM tends to become *stored* and *connected* in LTM (Clark et al., 2012; Dehaene, 2020, p. 160-161). WM is also a *bottleneck* during both problem solving and learning because storage in WM is very limited.

WM Limits

During problem solving, the slots of WM must hold the *novel* (new or problem specific) *chunks* of problem data. Cognitive science defines a *chunk* as a collection of encoded sensory (e.g., visual, sound) elements that have been stored and wired together in LTM and have meaning (Anderson, 1996; Dehaene, 2020, p. 6; Willingham, 2006). Novel data chunks that must be stored in WM slots include the problem goal, initial data/information, and answers found at middle-steps that must be looked up or mentally calculated rather than quickly recalled from LTM (Alloway, 2006; Luck & Vogel, 2013; Willingham, 2006).

WM is limited in capacity by its number of slots. During problem solving, if the data are sounds or symbols such as letters, words, or numbers. WM capacity is an average of 2 to 3 slots at age 7, gradually increasing to an adult average of four slots at age 14. WM capacity varies among individuals but is 3 to 5 slots for most adults. It remains essentially constant through adulthood until old age, when it has been found to decrease (Cowan, 2001, 2010; Gathercole, 1999, 2011).

WM is also limited in duration. Repeatedly reciting information orally or mentally can keep it stored in WM, but during other processing, including problem solving, each slot in WM can retain a chunk of novel data for only 30 seconds or less (Peterson & Peterson, 1959).

WM Strengths

If there were no way around WM limits, with only 3 to 5 limited-duration slots in WM, we could solve only very simple problems. But WM also has strengths. Studies have found a virtually unlimited number of facts and procedures can be applied to solving problems–if they are quickly recallable from an individual's LTM (Ericsson & Kintsch, 1995; Geary et al., 2008). Kirschner et al. (2006) summarize that for facts and procedures that have not previously been well-memorized, WM is "very limited in duration and in capacity," but "when dealing with previously learned information stored in long-term memory, these limitations disappear" (p. 77).

WM Implications

How does memorization circumvent WM limits? Try this brief experiment.

> Place a sheet of paper where you can write on it. Hold a pencil or pen.
>
> Now read silently three times in succession: 3.14159265
>
> Look away and write only the value 100 times larger.
>
> Then read silently three times in succession: 26494.1783
>
> Look away and write only the value 10 times larger.

Which problem was easier? You have previously memorized all or part of the first number as one 'chunk.' Holding that chunk takes one WM slot. Holding each number not memorized and applying the processing instruction took additional slots, but needing limited slots, your processing likely reached an answer close to correct.

The second problem required storage of a larger number of chunks of numerical data in WM slots, plus the slot or two for processing. The higher the number of slots needed, the more WM is likely to *overload* and lose stored information, leading to confusion and/or less success (Willingham, 2009c, p. 15).

Your ability to solve math problems in a topic depends in large measure on the number and size of chunks you have thoroughly memorized about the topic.

WM Summary

Neuroscience and cognitive psychology explain learning using varied terminology, but all current models for problem solving include space in WM that is stringently limited for novel data with few recallable relationships but essentially unlimited for relationships quickly recallable from LTM.

HOW THE BRAIN STORES LEARNING

The limits and strengths of working memory are one important discovery of recent cognitive research. How information is stored in and recalled from LTM is another. Sweller et al. (2010) note LTM "is not used to store random, isolated facts but rather to store huge complexes of closely integrated information that results in problem-solving skill" (p. 1304).

To learn is to *change LTM* in ways that either store and/or connect new information, or better connect and/or more strongly store previously stored information. The human brain has over 80 billion neurons. Each can be wired to about 10,000 other neurons via synaptic connections whose strength can vary (Dehaene, 2020, p. 10, 86-89). This structure gives human LTM an enormous potential capacity, but knowledge must be stored gradually. As one example, between infancy and college graduation, US students on average learn the meaning of about 17,000 'root words' at a rate of about 2-3 words per day (Biemiller, 2001).

Firing and Wiring

WM and LTM store sensory inputs, such as images and sounds, as encoded elements. For example, when first learning to recognize the symbol '5', a child by attention moves the symbol into WM, and with instructional guidance *processes* the shape by noting its component lines and curves (Anderson & Neely, 1996; Dehaene, 2020, p. 6, 86; Furst, 2020).

WM then uses each line and curve as a data *cue* to search LTM for a *match* (Dehaene, 2020, p. 90, 150; Ericsson & Kintsch, 1995). Because lines and curves are fundamentals stored early in LTM development, those WM cues find their match. LTM neurons storing the match are then said to *activate*: They fire (Furst, 2018, 2020). In neuroscience, a fundamental (but simplified) rule of learning is, 'neurons that *fire* together, *wire* together' (Hebb, 1949).

During repeated processing in WM, the visual image of '5' that becomes stored in LTM and its component elements in LTM fire at close to the same time, become wired together, and form a chunk of stored knowledge. With varied practice, different ways to represent '5' in fonts and hand-written become wired into a larger chunk of ways to represent the symbol '5.'

Chunking

With additional instruction and practice, a child stores and connects in LTM additional chunks of learning, including that '5' is spelled 'five' (if in English), is a number not a letter, represents a count of ••••• objects, and the concept that a number 'counts how many.'

After this larger chunk is wired into LTM (*memorized*), in a future problem, if 'five' or '5' or ••••• objects enters WM, it serves as a cue to search LTM for its match. If found, both the match and chunks wired to the match tend to activate (fire). Chunks previously retrieved together most often tend to be most quickly and strongly activated and most likely to be recalled by WM to solve the problem (Dehaene, 2020, p. 88). WM treats LTM chunks activated as a part of the single stored WM-data chunk. As a result, activated LTM chunks do not require additional slots in WM and do not increase the risk of WM overload.

More complex problem solving and learning follows similar steps. By the process of cued recall (or *cued retrieval*), data entering WM activates *prior learning* in LTM, which can be applied to solve the problem (Willingham, 2008). A chunk of problem data moved into a WM slot becomes a larger chunk, including *many* recallable relationships, if its match in LTM has many previously wired connections (Ericsson & Kintsch, 1995; Furst, 2020).

How Learning Happens

Neuroscience educator Efrat Furst (2018) writes that "processing in working memory is … information's 'entry ticket'" to LTM storage and retrieval. During problem solving, if a step is "perceived as successful" (Clark, 2010, p. 2), a record of any new data chunks plus other chunks activated in LTM is stored in what can be termed the brain's 'middle-term' memory (located in the brain's hippocampus) (Trafton, 2017). If similar problems are *not* solved again over several days, the initial ability to recall the new knowledge and its relationships to previously stored knowledge tend to be lost (Dehaene, 2020, p. 216; Willingham, 2015).

But if chunks are repeatedly processed together over multiple days, a record of the processing tends to become *consolidated*: related chunks that fire at close to the same time wire to form larger chunks in LTM (Furst, 2020; Trafton, 2017). As problems of increasing complexity are solved, *chunking* gradually wires into LTM larger, more strongly wired, and better-organized *conceptual frameworks* for topics (each termed a *schema*, plural *schemata*) (Kalyuga et al., 2003).

Using advanced microscopy, neuroscientists have imaged the growth of new connections among neurons in response to repeated successful problem solving. Imaging has also recorded the gradual decay of connections that have *not* been activated recently, which tends to result in *forgetting*. The brain's ability to learn and forget is termed its *plasticity* (Dehaene, 2020, p., 87, 137; Furst, 2019; Yang et al., 2014).

WORKING AROUND WM LIMITS

For a given topic, if LTM has many relevant chunks that can be activated by cues, substantial resources are available to solve problems. But what happens if the chunks a student has previously stored are few in number or limited in connections?

Avoiding Overload

At each step in problem-solving, if a needed relationship is not activated by a problem cue, it must be looked up and/or calculated. WM then needs to store in its slots both the data cue and the looked-up relationship. Unless the problem is very simple, the limited slots for data will tend to be already full, resulting in WM *overload* (Gathercole & Alloway, 2004; Willingham, 2004).

Furst (2018) writes, "overload leads to information loss–either incoming information will not be processed, or an 'in-process' item will be dropped for a new one." For example, in a multi-step problem, if solving 56/7=? is a required step and 56/7=8 has been memorized as a single chunk, one slot for the data and answer is needed. If a calculator or mental calculation is needed to answer, the 56 and seven and division operator require three WM slots and calculating then storing '8' requires at least one more slot. If the available slots have already been filled by other problem data, WM likely overloads (Willingham, 2009b).

When solving a problem, speed is also important. Using a calculator, mentally calculating, or searching the internet takes time. Because WM storage during processing is 30 or fewer seconds, a search or calculation tends to cause data being held in novel slots to 'time out' and be lost from WM, leading to confusion. In contrast, if a relationship is quickly recallable, activation that decays can quickly be reactivated by the data cues (Geary et al., 2008; Kirschner et al., 2006).

Automating Fundamentals

Students learn more quickly when instructors teach the strategies necessary to work around WM limits. Cognitive scientist Daniel Willingham has been a leader in efforts to inform educators of findings of research on learning. Willingham (2004) writes that the "lack of space in working memory is a fundamental bottleneck of human cognition" but identifies three strategies found by science that circumvent WM limits: chunking (discussed above), automaticity, and algorithms.

In mathematics, a 2008 report by cognitive experts Geary et al. (2008) advised educators:

> [T]here are several ways to improve the functional capacity of working memory. The most central of these is the achievement of automaticity, that is, the fast, implicit, and automatic retrieval of a fact or a procedure from long-term memory. (p. 4-5)

Let us examine the components of this central finding of cognitive research. Achieving recall with *automaticity* is also termed *automatization* or *automating* recall. For a fact such as six times seven, automated recall means the fact is so well memorized that "the answer is not calculated but simply retrieved from memory" (Willingham, 2009b, p. 16). If recall has been automated, slots in WM are not needed for processing, so the calculation and answer do not cause overload.

Automated recall of a *procedure* is retrieval of an algorithm: a sequence of steps, remembered as a single chunk, that successfully solves a certain type of problem.

Geary et al. (2008) add:

> [I]n support of complex problem solving, arithmetic facts and fundamental algorithms should be thoroughly mastered, and indeed, over-learned, rather than merely learned to a moderate degree of proficiency. (p. 4-6)

Overlearning is defined by cognitive science as "sustained practice, *beyond* the point of mastery" in recalling knowledge from LTM (Willingham, 2004).

Together, these findings mean that to work around WM limits, students must thoroughly memorize facts and procedures used frequently. The good news is cognitive science has also identified strategies that help students overlearn both facts and algorithms efficiently and effectively.

OVERLEARNING FACTS

Mathematical *facts* include wired chunks that precisely state a relationship among smaller chunks of knowledge (e.g., 14-8=6), or definitions such as "a right triangle has one interior 90^{o} angle," or rules such as the Pythagorean theorem.

Rehearsal

Automating factual recall begins with *rehearsal*. Repeating new information orally, mentally, or in writing is termed *maintenance rehearsal* (Belmont & Butterfield, 1971). To speed learning, maintenance rehearsal should involve as many senses as possible. Hearing, seeing, saying, and writing both words and symbols help to *multiply encode* a new fact into LTM (Paivio, 2014). MacLeod and Bodner (2017) found, relative to silent reading, "saying, writing, or typing" items yields substantial benefits in cued recall.

Elaborative rehearsal is thinking about meaning how new information is related to knowledge already retrievable. Elaborative rehearsal especially assists when facts are complex or likely to be encountered in specific contexts (Reder & Anderson, 1980). When learning math facts, elaborative rehearsal might involve manipulatives or number lines.

For simple facts such as '8+6=14', both types of rehearsal should be practiced, but to gain automated recall, maintenance rehearsal will be more efficient.

Retrieval Practice

Retrieval practice is effort to use one part of a relationship as a cue to recall other parts from LTM. Specific strategies include flashcards, cover/copy/compare, writing memorized algorithms (such as the quadratic formula or the mnemonic SohCahToa), or recalling sequences (2, 4, 8, 16, 32, …). To efficiently learn a significant number of unfamiliar verbatim facts, repeated retrieval practice is nearly always required (Carpenter & Agarwal, 2019; Powell, 2018).

Rehearsal should be sufficiently practiced before retrieval. Flashcards can assist in both. To rehearse, read the question, flip, read the answer. To retrieve, read, try to recall, flip. In his book *How We Learn*, neuroscientist Dehaene (2020) explains,

> The more you test yourself, the better you remember … [A]ctive engagement followed by error feedback maximizes learning … Using flashcards, try to remember the answer (prediction) before checking … (error feedback). (p. 186)

Retrieval practice applies the *testing effect*: a finding that both storage of information in and recall from LTM are improved when the learner practices retrieval. Cognitive studies have found recall is strengthened more by retrieval testing, including low stakes self-testing, than strategies such as highlighting or re-reading (Dunlosky, 2013; McDaniel et al., 2014).

Willingham (2008, p. 18) advises, “memories are formed as the residue of thought.” We tend to remember (learn) what we think about.

Distributed Practice

Retrieval practiced several times in one sitting (such as the night before a test) but not over multiple days is termed *massed practice* or *cramming*. Crammed knowledge can help on a test within a day or two (increasing *performance*) but tends to be quickly forgotten (it tends not to be *learned*) (Willingham, 2002, 2015).

Retrieval practiced over several days is termed *distributed practice* (Carpenter and Agarwal, 2019). This *spaced retrieval* tends to ‘flatten the forgetting curve,’ meaning what has been learned tends to be forgotten more slowly and, thus, better remembered for longer periods (Ebbinghaus, 1885/1913/2013). By providing intervening sleep, spaced practice promotes better memory *consolidation*: the improved organization of wired chunks within LTM (Dehaene, 2020, p. 221-230; Yang et al., 2014).

Spacing day or longer time gaps between retrieval-practice sessions leads to some forgetting, creating what is known as a *desirable difficulty*: The increased mental effort required for later retrieval (the difficulty) promotes longer lasting cued recall (which is desirable). Bjork and Bjork (2011) write,

> Many difficulties are undesirable during instruction and forever after. Desirable difficulties … are desirable because they trigger encoding and retrieval processes that support learning, comprehension, and remembering. (p. 62)

Overlearn With Spacing

Recall repeatedly practiced to perfection, repeated over days and on occasion in weeks and months thereafter, is termed *spaced overlearning***.** After spaced overlearning, forgetting may occur, but quick recall can generally be restored with far less re-study. Cognitive scientists call this the ‘savings in relearning’ effect (Willingham, 2002, 2015).

Which facts are a priority to overlearn? Willingham (2004) suggests, “Knowledge that will be used again and again.” During careers, specifics of work will change, but fundamentals overlearned with spacing will likely be recallable, with occasional ‘refreshing of memory,’ for a lifetime (Bjork & Bjork, 2019).

AUTOMATION OF ALGORITHMS

An *algorithm* (also termed a *well-structured* or *fixed procedure***)** is a ‘recipe’ that solves a type of complex problem in a fixed sequence of steps (Van Merriënboer & Kirschner, 2007; Willingham 2009b). A useful algorithm is one that, for a specific type of problem, has empirically proven to convert problem data step-by-step, without overloading WM at any step, to reach a correct answer (Geary et al., 2008; Willingham, 2009a, 2009b). Examples of algorithms in mathematics include sequences remembered by mnemonics (PEDMAS or PEMDAS), standard algorithms of arithmetic and algebra, and steps of a worked example.

Most (but not all) algorithms work by breaking complex problems into a series of smaller steps that can each be answered by applying memorized fundamentals. If the middle-step answers of an algorithm can be stored on a medium, such as a manipulative or paper and pencil, instead of in the mental slots of WM, overload is less likely. Examples of ‘non-mental storage on paper’ are seen in the standard algorithms of multi-digit arithmetic (Geary et al., 2008 p. 4-32).

Neuroscience and Algorithms

Dehaene (2020) describes WM as the brain’s “central bottleneck” during problem solving, but explains how neuroscience has found automated algorithms circumvent WM limits:

> Automatization mechanisms ‘compile’ the operations we use regularly into more efficient routines. They transferred them to other brain circuits, outside our conscious awareness … As long as a mental operation remains effortful because it has not yet been automated by overlearning, it absorbs valuable executive attention resources … Consolidation is essential because it makes our precious brain resources available for other purposes. (p. 222-223)

Implicit Retrieval

Working around WM limits requires the “fast, *implicit*, and automatic retrieval of a fact or a procedure” from LTM (Geary et al., 2008, p. 4-5). Science defines *implicit retrieval* as *intuitive* or *tacit*, which may not include a conscious ability to state what was recalled or why it was recalled (Clark, 2006, 2010). For example, take a moment to solve (try ‘in your head’ first): $3x+2=29$ for x.

Most likely, if you teach math (i.e., you are an expert), the visual structure intuitively activated a well-memorized algorithm. You quickly recalled steps and facts to reach $x=9$. You did not ‘explain why’ you subtracted two before dividing by three. That’s important because if a procedure is part of a more complex problem, by solving quickly, stored problem data is less likely to ‘time out’ of WM.

Mathematicians and math educators must be able to explicitly *explain why* they do steps. When first learning a procedure, students aided by a teacher (by *modelling*) need to be able to explain how it works. But for the vast majority of students, after initial learning, explicitly ‘explaining why’ takes time and is not necessary. Most students need math as a *tool* to solve problems in other complex fields. When using a tool, the goal is to automate: to carry out the right steps quickly without conscious attention to ‘why.’ Automated recall is often subconscious (Clark, 2006, 2010).

As one measure of relative need, US colleges graduate about 25,000 math majors each year (Trapani & Hale, 2019) but millions enter careers in which automated math is needed as a *tool* to solve problems. While they are students, those millions need time to practice learning and applying facts and procedures. With that practice, they gain the implicit understanding needed to solve math problems using facts and steps recalled intuitively (Clark 2006, 2010).

Retrieval Cues

Out of hundreds of algorithms that students learn in K-12 math, how does the brain recall which algorithm to apply to a specific type of problem? As the steps, typical cues, and typical contexts for an algorithm's use are practiced when solving a topic's problems, all become wired into a larger chunk in LTM. After a topic's problems have been practiced in a variety of distinctive contexts, the brain automatically notes a new problem's components and retrieves an appropriate algorithm to apply–in a manner similar to how the brain intuitively chooses an appropriate phrase to express an idea (Clark, 2006, 2010; Pinker, 2007).

When teaching a new algorithm, cognitive studies recommend including both different typical retrieval cues and distinctive topic contexts to help the brain sense when the algorithm is needed (Willingham, 2003, 2008). Anderson and Neely (1996) write,

> Retrieval cues can be anything from components of the desired memory to incidental concepts associated with that item during its processing … Retrieving a target item [occurs] when the cues available at the time of recall are sufficiently related to that target to identify it uniquely in memory. (p. 239)

Keeping Slots Open

To be useful, an algorithm must work whether data are single digits that can be solved by mental math or multi-digit solved with a calculator. But learning a new algorithm is accelerated if examples and initial calculations are contrived to be solved using overlearned mental arithmetic. Why?

When practicing the use of an algorithm, as more steps are automated, more slots in WM become available to hold and process (and thereby learn) the sequence. If sample problems can be solved using automated mental math, more slots are available to hold and process the context cues that differentiate problem types. By designing sample problems to keep WM slots open, new algorithms and appropriate contexts to implicitly apply them are learned more quickly.

Interleaved Practice

Assignments that repeat problems of the same type are termed *blocked practice.* Mixing different types within a problem set is *interleaved practice* (Taylor & Rohrer, 2010). After initial blocked practice of a new algorithm, interleaved practice that follows helps the learner explicitly and implicitly spot differences amongst similar problems (i.e., discrimination learning), and spot/remember similarities among different problems. Using interleaving, solving is initially more difficult, but the difficulty is desirable because long-term, students will better remember the cues and contexts associated with each algorithm (Bjork & Bjork, 2019).

Standard Algorithms

For each problem type, many algorithms will likely work, but for students, cognitive studies support initial mastery of one 'standard' algorithm for each problem type. If multiple algorithms are practiced before the standard is overlearned, choosing an effective algorithm to apply is less likely to become intuitive. Further, practicing several algorithms that are similar but not the same leads to cognitive 'interference' when attempting to remember which steps and sequence comprise each procedure (Anderson & Neely, 1996; Dewar et al., 2007).

CONCEPTS AND REASONING

Knowledge in mathematics can be categorized as factual (declarative), procedural (algorithmic), or conceptual. Geary et al. (2008, p. 4-7) write that "conceptual knowledge refers to general knowledge and understanding" stored in LTM. An example of a concept is the commutative property of addition: (a+b=b+a).

Cognitive studies emphasize conceptual understanding is vitally important and needs to be taught (Siegler & Lortie-Forgues, 2015). Conceptual understanding helps "facilitate transfer and long-term retention" of knowledge, organizing knowledge within LTM by the deeper structure of its meaning (Geary et al., 2008, p. 4-7). After initial consolidation (wiring of chunks), during additional topic learning, conceptual understanding helps as LTM *re-consolidates*: re-wires to better organize topic schema (Furst, 2020). Repeated re-consolidation leads to "a shift from slow, conscious, and effortful processing to fast, unconscious, and automatic expertise" (Dehaene, 2020, p. 222).

Facts Before Concepts

However, independent of how important concepts are, they need to be preceded by the learning of facts upon which the concepts are based. Willingham (2009b) writes that among facts, procedures, and concepts, "conceptual knowledge is the most difficult to acquire … [A] teacher cannot pour concepts directly into students' heads. Rather, new concepts must build on something students already know" (p. 18).

When teaching concepts, Willingham (2002) notes that if examples have components that are *concrete* and *familiar* (well-memorized), slots in WM are more likely to be available to hold conceptual elements during processing. Willingham (2009b) suggests instructors should "[E]xplain to students that automaticity in facts is important because it frees their minds to think about concepts" (p. 19).

Reasoning–or Remembering?

For students, without reliance on a memorized algorithm, reasoning based on concepts may work to solve very simple problems. However, in a math problem with multiple steps, 'general reasoning' is nearly always too general to provide reliable cognitive guidance on which steps to solve the problem will be successful (Sweller et al., 2010). For novice learners, trying to evaluate different reasoning approaches nearly always overloads WM (Kirschner et al., 2006).

The non-expert brain is strong at remembering, not reasoning. Anderson (1996), a leader in cognitive research, advises:

> One fundamentally learns to solve problems by mimicking examples of solutions … [A]cquiring competence is very much a labor-intensive business in which one must acquire one-by-one all the knowledge components … We need to recognize and respect the effort that goes into acquiring competence. (1996, p. 359)

Willingham (2009c, p. 68-72) states as the cognitive principle, "the brain was not designed for thinking." Willingham (2009a) notes, "much of the time that we see people apparently engaged in logical thinking, they are actually engaged in memory retrieval" (p. 8). When students seem to solve complex problems based on "understanding" without a recalled algorithm, research nearly always finds "understanding is remembering in disguise" (Willingham, 2009c, p. 70).

STANDARDS FOR ARITHMETIC FACTS

In 2010-11, over 40 US states adopted the CCMS to govern K-12 instruction. As in the 1989 NCTM standards, the CCMS state as a fundamental assumption that students can "reason abstractly" to solve problems (NGA and CCSSO, 2010, p. 6). The CCMS make no mention of what science has discovered regarding WM limits, automaticity, distributed retrieval practice, or reasoning. Among those who drafted and voted to approve the 2010 CCMS, none included a scientist who studies how the brain solves problems (NGA, 2010; Zimba, 2014).

In preparation for this paper, math standards were examined in all U.S. states as of January 2020. This search found that since 2014, as reported elsewhere (Heitin, 2015; Norton et al., 2017), a number of states have dropped the 'original CCMS' as their adopted standards, but in over 30 states, math standards remained either the same as, or very similar to, the 2010 CCMS.

Cognitive research can help in designing effective standards, but when new science is discovered, assumptions that are the basis of current standards must be reviewed. How do the 2010-type standards still in use in most states compare to strategies informed by cognitive science in 2023?

Grade 1 and 2

In over 30 states in 2020, standards for learning addition and subtraction facts were either the same as or essentially identical to the following two CCMS standards (the standard's first number is the grade level).

1. **1.OA.6.** Add and subtract within 20, demonstrating fluency for addition and subtraction within 10. Use strategies such as counting on; making 10 (e.g., 8+6=8+2+4=10+4=14); decomposing a number leading to a ten (e.g., 13-4=13-3-1=10-1=9); using the relationship between addition and subtraction (e.g., knowing that 8+4=12, one knows 12-8=4); and creating equivalent but easier or known sums (e.g., adding 6+7 by creating the known equivalent (6+6+1=12+1=13).
2. **2.OA.2**. Fluently add and subtract within 20 using mental strategies [from **1.OA.6**]. By the end of grade 2, know from memory all sums of two one-digit numbers. (NGA and CCSSO, 2010)

In a 2016 interview, one of three 'lead writers' of the CCMS was asked if in the standards, "being fluent" means "knowing from memory." He answered:

> "They are not the same thing, and the language of the standards makes this clear … [In **2.OA.2**], if being fluent were the same thing as knowing from memory, the second sentence would not have been necessary … In every case, fluency pertains to an act of calculation." (Northern, 2016)

To summarize: CCMS and most states ask students in grades 1 and 2 to solve the basic addition and subtraction fundamentals by "fluent calculation" using the "*mental* strategies" (**2.OA.2**) in standard **1.OA.6**. These standards raise multiple concerns.

First, most of the **1.OA.6** strategies call on students to mentally solve calculations in ways science says the brains of most cannot. At ages 6-8, children have especially limited WM capacity. To add single-digits, grade 1-2 students may be able to use the memorized algorithm that is counting to mentally manage "counting on."

But to solve the more complex **1.OA.6** calculations mentally, a student would need to move into WM the problem goal (add or subtract?) and its two data numbers, requiring three WM slots. Reasoning to choose a strategy to use would take additional slots. Applying the strategy may also require slots. Cowan's (2001, 2010) measurements found for grade 1-2 students, WM capacity averages only 2-3 slots.

Gathercole (2008) notes that "poor working memory skills are relatively commonplace during childhood" and advises to "restructure complex tasks" to "avoid working memory overload" (p. 384).

Facts recallable from memory do not contribute to overload, but the CCMS and similar standards do not call for any basic addition facts to be known "from memory" until "the end of grade 2." How will they solve the complex **1.OA.6** mental calculations during grade 1?

A second concern: The complex strategies required in the grade 1-2 standards of most states are unlikely to result in fundamentals becoming recallable. Kirschner and Hendrick (2020) emphasize "children are not 'little adults.'" In grade 1-2, students lack the combination of WM capacity and knowledge that adults possess that is needed to solve multi-step calculations.

What students *can* do in grades 1-2 is to make math facts recallable using flashcard-type strategies. Recallable facts will be needed to avoid overload when learning to solve more complex problems in higher grades. Science says "the most effective" strategy to make facts quickly recallable is distributed retrieval practice (Dunlosky, 2013).

Starting from infancy, retrieval with feedback can achieve automated recall (Fazio & Marsh, 2019; Fritz et al., 2007). Burns et al. (2019) found the combination of rehearsal and retrieval practice to be both efficient and effective in the elementary grades. For teachers, a measurement that a fact has been automated, not calculated, is recall within 2 to 3 seconds (Burns et al., 2014).

During childhood, learning spoken *phrases* is powerfully instinctive. With practice, most kindergarteners can recite lengthy word sequences (counting to 100, book passages, songs). If a first-grader practices reciting "eight plus six is fourteen" repeatedly over several days, it becomes one of thousands of phrases they can complete automatically (MacLeod & Bodner, 2017; Pinker, 2007).

Manipulatives can help with calculations, but their use takes time and WM slots. Strategies less efficient than spaced retrieval will result in knowing some facts eventually, but time for math is limited in grade 1-2 classrooms. Until recall of all arithmetic basics is automated, understanding examples to learn higher-level calculations will be limited.

A third concern: Most states do not ask that any *subtraction* facts be known "from memory" at any point, asking instead that subtraction facts be "fluently" calculated "using the relationship between addition and subtraction" (**1.OA.6**).

During grade 1, most states ask that students calculate addition facts, rather than know them from memory. Asking grade 1 students to first calculate the addition fact, then calculate the subtraction fact, given grade 1 WM capacity, is likely to cause overload.

In higher grades, if the paired addition fact is by then recallable, using the inverse relationship to find a simple subtraction fact may work, but the calculation requires a WM slot to hold the answer. To learn higher-level algorithms and concepts efficiently, the strategy must be to automate as many components of problem solving as possible. WM slots will be needed to process step sequences and conceptual elements. For simple facts, calculation to subtract may make sense mathematically, but to prepare for higher-level learning, it is the wrong strategy cognitively.

Summarizing cognitive research, Willingham (2009b) writes that for all of the basic facts

> of addition, subtraction, multiplication and division, … answers must be well learned so that … the answer is not calculated but simply retrieved from memory, … [A]utomatic retrieval of basic math facts is critical to solving complex problems … Calculating simple arithmetic facts does indeed require working memory. (2009b, p. 16-17)

Concepts

Students need implicit understanding of concepts, facts, and procedures, but "new concepts must build on something students already know" (Willingham, 2009b).

To begin to learn mathematics, by rehearsal and retrieval, students 'rote memorize' how to count. The concept of addition and subtraction can then be learned from simple examples of "counting on" (**1.OA.6**) and 'counting back' on fingers using numbers 10 and under.

To teach place value, simple examples of other 1.OA.6 calculations may be manageable on paper or using manipulatives. But learning concepts by solving complex calculations *mentally*, as called for in 2.OA.2, is not likely to be reliably successful for most students until they are older, when WM capacity has increased, and basic facts have been overlearned.

Grade 3 and 4

The CCMS and most states call for learning the over 120 verbatim, non-trivial *multiplication* and *division* facts in a manner similar to the addition and subtraction facts. They do not ask that any division facts be recallable. They call for learning the multiplication facts 'from memory' using calculation strategies far too inefficient to automate recall in the time available for math in elementary classrooms.

The foundations for mathematics are the over 120 non-trivial addition and subtraction and over 120 multiplication and division facts. Automaticity in recall of all basic facts must be a priority in grades 1-4. Distributed retrieval practice is the most efficient way to achieve automaticity. Even with this efficiency, standards should make clear: Substantial class time must be budgeted to not just learn but overlearn these over 240 verbatim fundamentals.

To learn the procedures for *multi-digit* arithmetic, in grades 3-4 the CCMS and most states call for a year or more practicing non-standard algorithms before learning the standard algorithms. Because of cognitive 'interference,' science-informed standards would call for automating a standard algorithm for each operation before practicing non-standard procedures.

Non-Typical States

From two states, there is good news. *TIMSS* is a test given every four years to 4th and 8th graders in about 50 nations including the US. In 2011 results, evaluating 8th grade "number" skills, top scoring nations in order were South Korea, Singapore, and Japan.

Subdivisions of nations may also ask to have scores reported. Among US states that did so, Massachusetts scored 567, above Japan's 557 (MDESE, 2017). Minnesota scored 556. The US ranked poorly overall with a score of 514 (Hartman & Nelson, 2016; NCES, 2015; Provasnik, 2012).

What accounts for the high scores from Massachusetts and Minnesota? One factor may be, unlike most states, in most years from 2000 to 2020, Massachusetts required memorization of all basic math facts (MDESA, 2017; Riordan & Noyce, 2001). Minnesota did not adopt the CCMS and has had a policy of 'no calculators' on state elementary tests since 2003, encouraging the overlearning of arithmetic fundamentals (MDE, 2003, 2019).

CONCLUSIONS

Science has discovered that when the brain tries to reason with not-well-memorized information, stringent limits apply. In publications for educators, the importance of automaticity to work around WM limits has been noted since 1996 (Hirsch, 1996). Yet since that time in most states, key K-12 standards have continued to require young students to solve math problems by reasoning in ways science says their brains simply cannot manage.

Why has the progress of science been ignored? In every US state, political officials decide K-12 standards (Loveless, 2021). In most states, those officials have been lobbied, often with success, by education-theorists who philosophically oppose the repeated effort ('drill') that cognitive science says is required to circumvent the brain's working memory limits. As test scores indicate, resulting standards have not been effective.

In this article, our review of US standards has been limited to fundamentals of arithmetic. Our analysis is not intended to be the final word, and we invite responses. But the authors posit that these data show in most states, K-12 standards would benefit from a review–with participation by scientists who study learning. In medicine, standards are set by practicing doctors guided by science. In teaching, the most effective standards will be set by practicing front-line educators guided by science.

If the science of learning becomes a more significant factor in mathematics standards and curricula, student learning will gradually improve. But to prepare for careers in the sciences, engineering, and technology, students need 17 years under effective standards–in grades K-12, then college. When will those who decide standards put the progress of science to good use? How long can the US afford to wait?

Author contributions: All authors have sufficiently contributed to the study and agreed with the results and conclusions.

Funding: No funding source is reported for this study.

Ethical statement: The authors stated that for this review of research, no ethics application was required.

Declaration of interest: No conflict of interest is declared by authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the corresponding author.

REFERENCES

Alloway, T. P. (2006). How does working memory work in the classroom? *Educational Research and Reviews, 1*(4), 134-139.

Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem situations. *Psychological Review, 94*(2), 192-210. https://doi.org/10.1037/0033-295X.94.2.192

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist, 51*(4), 355. https://doi.org/10.1037/0003-066X.51.4.355

Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. L. Bjork, & R.A. Bjork (Eds.), *Memory* (pp. 237-313). Academic Press. https://doi.org/10.1016/B978-012102570-0/50010-0

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence, & J. T. Spence (Eds.), *Psychology of learning and motivation* (pp. 89-195). Academic Press. https://doi.org/10.1016/S0079-7421(08)60422-3

Belmont, J. M., & Butterfield, E. C. (1971). Learning strategies as determinants of memory deficiencies. *Cognitive Psychology, 2*(4), 411-420. https://doi.org/10.1016/0010-0285(71)90024-7

Biemiller, A. (2001). Teaching vocabulary. *American Educator, 25*(1), 24-28.

Bjork, R. A., & Bjork, E. L. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. E . Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world* (pp. 59-68).

Bjork, R. A., & Bjork, E. L. (2019). Forgetting as the friend of learning: Implications for teaching and self-regulated learning. *Advances in Physiology Education, 43*(2), 164-167. https://doi.org/10.1152/advan.00001.2019

Burns, M. K., Aguilar, L. N., Young, H., Preast, J. L., Taylor, C. N., & Walsh, A. D. (2019). Comparing the effects of incremental rehearsal and traditional drill on retention of mathematics facts and predicting the effects with memory. *School Psychology*, *34*(5), 521. https://doi.org/10.1037/spq0000312

Burns, M. K., Ysseldyke, J., Nelson, P. M., & Kanive, R. (2014). Number of repetitions required to retain single-digit multiplication math facts for elementary students. *School Psychology Quarterly, 30*(3), 398. https://doi.org/10.1037/spq0000097

Carpenter, S. K., & Agarwal, P. K. (2019). *How to use spaced retrieval practice to boost learning.* http://pdf.retrievalpractice.org/SpacingGuide.pdf

Clark, R. E., Kirschner, P. A., & Sweller, J. (2012). Putting students on the path to learning: The case for fully guided instruction. *American Educator, 36*(1), 5-11.

Clark, R. E. (2006). Not knowing what we don't know: Reframing the importance of automated knowledge for educational research. In G. Clareboutand, & J. Elen (Eds.), *Avoiding simplicity, confronting complexity* (pp. 1-14). Brill Sense. https://doi.org/10.1163/9789087901189_002

Clark, R. E. (2010). Cognitive and neuroscience research on learning and instruction: Recent insights about the impact of non-conscious. In *Proceedings of the International Conference on Education Research, Seoul*.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87-114. https://doi.org/10.1017/S0140525X01003922

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science, 19*(1), 51-57. https://doi.org/10.1177/0963721409359277

de Groot, A. D. (1946/2014). Thought and choice in chess. In A. de Groot, G. W. Baylor, & S. Sloan (Eds.), *Thought and choice in chess*. De Gruyter Mouton. https://doi.org/10.1515/9783110800647

Dehaene, S. (2020). *How we learn: Why brains learn better than any machine... for now*. Viking.

Delazer, M., Domahs, F., Bartha, L., Brenneis, C., Locky, A., & Trieb, T. (2004). The acquisition of arithmetic knowledge: An fMRI study. *Cortex, 40*(1), 166-167. https://doi.org/10.1016/s0010-9452(08)70936-5

Dewar, M., Cowan, N., & Della Sala, S. (2007). Forgetting due to retroactive interference. *Cortex, 43*(5), 616-634. https://doi.org/10.1016/S0010-9452(08)70492-1

Dhond, R. P., Marinkovic, K., Dale, A. M., Witzel, T., & Halgren, E. (2003). Spatiotemporal maps of past-tense verb inflection. *Neuroimage, 19*(1), 91-100. https://doi.org/10.1016/S1053-8119(03)00047-8

Dunlosky, J. (2013). Strengthening the student toolbox: Study strategies to boost learning. *American Educator, 37*(3), 12-21.

Ebbinghaus, H. (1885/1913/2013). Memory: A contribution to experimental psychology. *Annals of Neurosciences, 20*(4), 155. https://doi.org/10.5214/ans.0972.7531.200408

Ericsson, K. A., & Kintsch, K. (1995). Long-term working memory. *Psychological Review, 102*(2), 211-245. https://doi.org/10.1037/0033-295X.102.2.211

Fazio, L. K., & Marsh, E. J. (2019). Retrieval-based learning in children. *Current Directions in Psychological Science, 28*(2), 111-116. https://doi.org/10.1177/0963721418806673

Fritz, C. O., Morris, P. E., Nolan, D., & Singleton, J. (2007). Expanding retrieval practice: An effective aid to preschool children's learning. *Quarterly Journal of Experimental Psychology, 60*(7), 991-1004. https://doi.org/10.1080/17470210600823595

Furst, E. (2018). *Learning in the brain.* https://sites.google.com/view/efratfurst/learning-in-the-brain

Furst, E. (2019). *Understanding understanding*. https://sites.google.com/view/efratfurst/understanding-understanding

Furst, E. (2020). *Reconsolidation*. https://sites.google.com/view/efratfurst/reconsolidation

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences, 3*(11), 410-419. https://doi.org/10.1016/S1364-6613(99)01388-1

Gathercole, S. E. (2008). Working memory in the classroom. *The Psychologist, 21*, 5. https://doi.org/10.1016/B978-012370509-9.00179-0

Gathercole, S. E. (2011). Working memory: A cognitive system that supports learning? *Ross Trust Seminar*. https://www.rch.org.au/uploadedFiles/Main/Content/ccch/RT_Seminar_Susan_Gathercole.pdf

Gathercole, S. E. (2015). *Working memory skills and educational attainment*. https://www.slideshare.net/Swiss_Cottage/keynote-susan-gathercole

Gathercole, S. E., & Alloway, T. P. (2004). Working memory and classroom learning. *Dyslexia Review, 15*, 4-9.

Geary, D. C. (2002). Principles of evolutionary educational psychology. *Learning and Individual Differences, 12*(4), 317-345. https://doi.org/10.1016/S1041-6080(02)00046-8

Geary, D. C., & Berch, D. B. (2016). Evolution and children's cognitive and academic development. In D. C. Geary, & D. B. Berch (Eds.), *Evolutionary perspectives on child development and education* (pp. 217-249). Springer. https://doi.org/10.1007/978-3-319-29986-0_9

Geary, D. C., Boykin, A. W., Embretson, S., Reyna, V., Siegler, R., Berch, D. B., & Graban, J. (2008). Chapter 4: The report of the task group on learning processes, 4-2 to 4-10. *US Department of Education*. https://files.eric.ed.gov/fulltext/ED502980.pdf

Goodman, M., Sands, A., & Coley, R. (2015). *America's skills challenge: Millennials and the future.* Educational Testing Service.

Hartman, J., & Nelson E. (2016). *Automaticity in computation and student success in introductory physical science courses*. http://arxiv.org/abs/1608.05006

Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition, 177*, 263-277. https://doi.org/10.1016/j.cognition.2018.04.007

Hebb, D. O. (1949). *The organization of behavior*. Psychology Press.

Heitin, L. (2015). Nebraska's new math standards. *Education Week*. https://www.edweek.org/teaching-learning/nebraskas-new-math-standards-a-comparison-to-the-common-core/2015/09

Hirsch, Jr, E. D. (1996). *The schools we need: And why we don't have them*. Anchor. https://doi.org/10.2307/358685

Jonassen, D. H. (1997). Instructional design models for well-structured and ill-structured problem-solving learning outcomes. *Educational Technology Research and Development, 45*(1), 65-94. https://doi.org/10.1007/BF02299613

Kalyuga, S., Ayres, P. L., Sweller, J., & Chandler, P. (2003). The expertise reversal effect. *Educational Psychologist, 38*, 23. https://doi.org/10.1207/S15326985EP3801_4

King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgment*. Jossey-Bass.

Kirschner, P. A., & Hendrick, C. (2020). *How learning happens: Seminal works in educational psychology and what they mean in practice*. Routledge. https://doi.org/10.4324/9780429061523

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem based, experiential and inquiry-based teaching. *Educational Psychologist, 41*, 75-86. https://doi.org/10.1207/s15326985ep4102_1

Klein, D., Braams, B., Parker, T., Quirk, W., Schmid, W., & Wilson, W. (2005). *The state of state math standards, 2005*. Fordham Institute.

Loveless, T. (2003). The regulation of teaching and learning. In *Stability and change in American education: Structure, process, and outcomes* (pp. 171-192). Eliot Werner Publications.

Loveless, T. (2021). *Between the state and the schoolhouse: Understanding the failure of Common Core*. Harvard Education Press.

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity. *Trends in Cognitive Sciences, 17*(8), 391-400. https://doi.org/10.1016/j.tics.2013.06.006

MacLeod, C. M., & Bodner, G. E. (2017). The production effect in memory. *Current Directions in Psychological Science, 26*(4), 390-395. https://doi.org/10.1177/0963721417691356

Mahnken, K. (2021). Long-term NAEP scores for 13-year-olds drop for first time since testing began in 1970s. *The 74*. https://www.the74million.org/article/naep-long-term-unprecedented-performance-drop-american-13-year-olds/

McDaniel, M. A., Brown, P. C., & Roediger, H. L. (2014). *Make it stick: The science of successful learning*. Harvard University Press.

MDE. (2003). *Minnesota academic standards–Mathematics K-12: Grade 6*. Minnesota Deptartment of Education.

MDE. (2019). *Minnesota comprehensive assessments–Series III (2019)*. https://education.mn.gov/MDE/dse/test/spec/

MDESE. (2017). Massachusetts curriculum framework for mathematics-2017. *Massachusetts Department of Elementary and Secondary Education*. https://www.doe.mass.edu/bese/docs/fy2011/2010-12/MACurricululmFramework.pdf

NCES. (2015). Trends in international mathematics and sciences study (TIMSS). *National Center for Education Statistics*. http://nces.ed.gov/timss/benchmark.asp

NCES. (2021). *NAEP LTT*. https://www.nationsreportcard.gov/ltt/

NCTM. (1989). *Curriculum and Evaluation Standards for School Mathematics*. National Council of Teachers of Mathematics.

NGA and CCSSO. (2010). *Common Core state standards initiative: Common Core state standards for mathematics*. National Governors Association and the Council of Chief State School Officers.

NGA. (2010). Common Core state standards initiative: K-12 standards development teams. *National Governors Association*. http://www.regents.la.gov/assets/docs/2013/09/State-Educators-Involvement-in-Development-of-CCSS.pdf

Northern, A.M. (2016). *Does Common Core math expect memorization?* Fordham Institute. https://edexcellence.net/articles/does-common-core-math-expect-memorization-a-candid-conversation-with-jason-zimba

Norton, J., Ash, J., & Ballinger, S. (2017). *Common Core revisions: What are states really changing*? Abt Associates.

OECD. (2012). *Literacy, numeracy and problem solving in technology-rich environments*. OECD Publishing.

Paivio, A. (2014). *Mind and its evolution: A dual coding theoretical approach*. Psychology Press. https://doi.org/10.4324/9781315785233

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58*(3), 193-198. https://doi.org/10.1037/h0049234

Pinker, S. A. (2007). *The language instinct: How the mind creates language*. HarperCollins.

Powell, S. R. (2018). *Fluency practice*. https://www.youtube.com/watch?v=86NZvtL640I

Provasnik, S. (2012). *Highlights from TIMSS 2011.* NCES 2013-009. National Center for Education Statistics.

Raimi, R. A., & Braden, L. S. (1998). *State mathematics standards*. Fordham Institute.

Reder, L. M., & Anderson, J. R. (1980). A comparison of texts and their summaries. *Journal of Verbal Learning and Verbal Behavior, 19*(2), 121-134. https://doi.org/10.1016/S0022-5371(80)90122-X

Riordan, J. E., & Noyce, P. E. (2001). The impact of two standards-based mathematics curricula on student achievement in Massachusetts. *Journal for Research in Mathematics Education, 32*(4), 368-398. https://doi.org/10.2307/749700

Seidenberg, M. (2017). *Language at the speed of sight*. Basic Books.

Siegler, R. S., & Lortie-Forgues, H. (2015). Conceptual knowledge of fraction arithmetic. *Journal of Educational Psychology, 107*(3), 909. https://doi.org/10.1037/edu0000025

Sweller, J. (2008). Instructional implications of David C. Geary's evolutionary educational psychology. *Educational Psychologist, 43*(4), 214-216. https://doi.org/10.1080/00461520802392208

Sweller, J., Clark, R., & Kirschner, P. (2010). Teaching general problem-solving skills is not a substitute for, or a viable addition to, teaching mathematics. *Notices of the American Mathematical Society, 57*(10), 1303-1304.

Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology, 24*(6), 837-848. https://doi.org/10.1002/acp.1598

Trafton, A. (2017). *Neuroscientists identify brain circuit necessary for memory formation*. http://news.mit.edu/2017/neuroscientists-identify-brain-circuit-necessary-memory-formation-0406

Trapani, J., & Hale, K. (2019). Higher education in science and engineering. In *Science & engineering indicators 2020.* National Science Foundation.

Van Merriënboer, J. G., & Kirschner, P. A. (2007). *Ten steps to complex learning*. Lawrence Erlbaum Associates.

Willingham, D. T. (2002). Allocating student study time. *American Educator, 26*(2), 37-39.

Willingham, D. T. (2003). Students remember what they think about. *American Educator, 27*(2), 37-41.

Willingham, D. T. (2004). Practice makes perfect, but only if you practice beyond the point of perfection. *American Educator, 28*(1), 31-33.

Willingham, D. T. (2006). How knowledge helps. *American Educator, 30*(1), 30. https://doi.org/10.12968/ftse.2006.5.4.20807

Willingham, D. T. (2008). What will improve a student's memory? *American Educator, 32*(4), 17-25.

Willingham, D. T. (2009a). Why don't students like school? *American Educator, 33*(2), 4-13

Willingham, D. T. (2009b). Is it true that some people just can't do math? *American Educator, 33*(4), 14-19.

Willingham, D. T. (2009c). *Why don't students like school?* Wiley. https://doi.org/10.1002/9781118269527

Willingham, D. T. (2015). Do students remember what they learn in school? *American Educator, 39*(3), 33-38.

Yang, G., Lai, C.S., Cichon, J., Ma, L., Li, W., & Gan, W. B. (2014). Sleep promotes branch-specific formation of dendritic spines after learning. *Science, 344*(6188), 1173-1178. https://doi.org/10.1126/science.1249098

Zimba, J. (2014). The development and design of the Common Core state standards for mathematics. *New England Journal of Public Policy, 26*(1), 10.

APPENDIX: DESIGNING MATHEMATICS STANDARDS IN AGREEMENT WITH SCIENCE

Working Memory Capacity

Cognitive studies find that the WM capacity of students is highly dependent on their age.

Experiments by Cowan (2001) found that when allowed to focus on ("attend to") a list of digits, adults could on average remember four to six, depending on the length (span) of the list, but when the digits were "ignored" to carry out steps of other cognitive processing, *adult* WM capacity averaged "4 ± 1."

Cowan's (2001) measurements also found 4th graders had lower WM capacity than adults, and first graders even less: a WM capacity during processing of only two-three digits (**Figure A1**).

In the case of some calculations, Cowan (2010) reported an average adult WM capacity of three, and WM capacity of seven year olds averaging 1.5.

Award-winning UK psychologist Gathercole (1999) has extensively studied working memory capacity in children. In tests of different components of working memory, Gathercole (1999) found relative increases in WM with age (**Figure A2**).

In another test, Gathercole (1999) measured relative WM capacity, rather than Cowan's (2010) measurements of absolute WM capacity, based on a recall test, the Working Memory Test Battery for Children (WMTB-C) (Gathercole, 2011, 2015; Gathercole & Alloway, 2004). Her results can be found in **Figure A3**:

1. WM capacity increases gradually after age five, then plateaus at about age 12.
2. Six-year olds average about half the WM capacity of 12-year olds.
3. WM capacity at each age varies widely among individuals. At age eight, the WM capacity of the 10th percentile have the same capacity as the average 5-year old, and at 90th percentile have same capacity of the average 12-year old (red arrows, **Figure A3**).
4. For the age six cohort (typically first-graders), those at the 10th percentile have half the WM capacity of the 90th percentile.

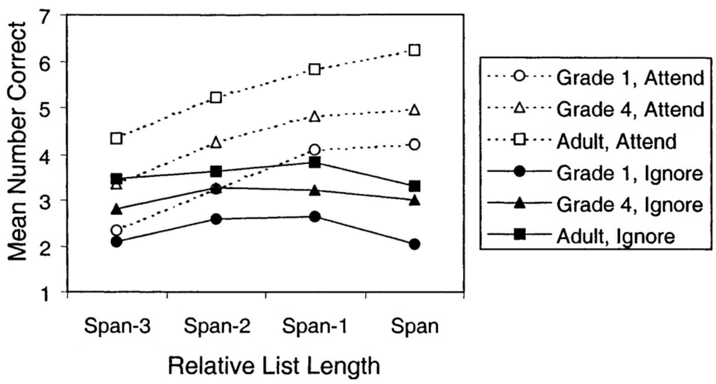

Figure A1. Measures of working memory capacity by age (Cowan, 2001)

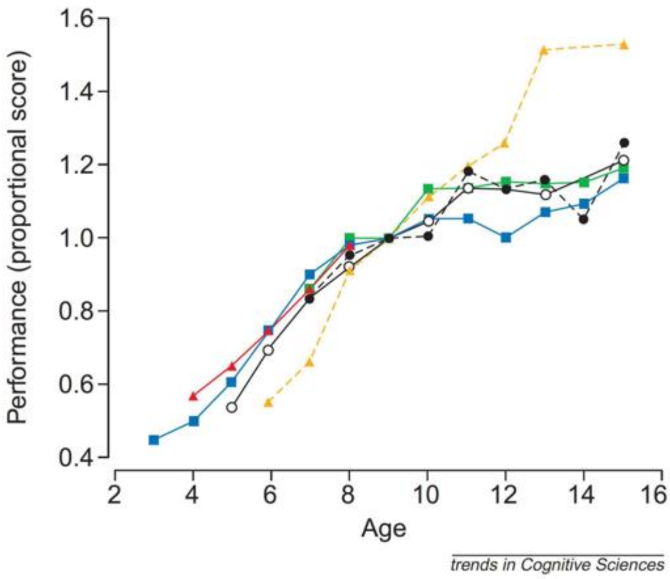

Figure A2. Performance on measures of relative working memory capacity by age (Gathercole, 1999)

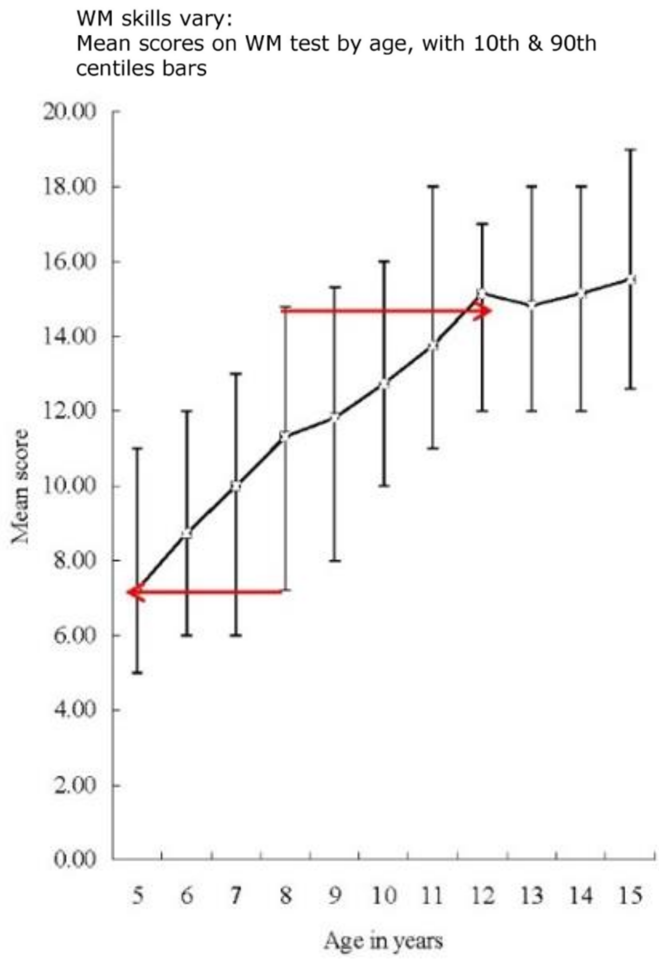

Figure A3. Mean scores on listening recall test from Working Memory Test Battery for Children (WMTB-C) as a function of age, with 10th & & 90th percentiles (Gathercole, 2011, 2015)