

This is the peer reviewed version of the following article:

Hobbiss, M., Sims, S., & Allen, R. (2020). Habit formation limits growth in teacher effectiveness: A review of converging evidence from neuroscience and social science. *Review of Education*, rev3.3226.

, which has been published in final form at <https://doi.org/10.1002/rev3.3226>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

Habit formation limits growth in teacher effectiveness: converging evidence from neuroscience and social science

Sam Sims^{a,*}, Michael Hobbiss^b, Rebecca Allen^a

^a *UCL, Institute of Education, UK*

^b *UCL, Institute of Cognitive Neuroscience, UK*

Teachers become rapidly more effective during the early years of their career but tend to improve increasingly slowly thereafter. This article reviews and synthesises converging evidence from neuroscience, psychology, economics and education suggesting that teachers' rate of growth slows because their practice becomes habitual. First, we review evidence suggesting that teaching is highly conducive to habit formation and that teachers display characteristic features of habitual behaviour. Next, we review empirical findings that performance asymptotes, as seen in teachers' learning curves, coincide with the reallocation of behaviour regulation to neural circuits governing habitual behaviour. Finally, original data is presented showing that teachers' behaviour becomes automatic around the time that many teacher effectiveness begins to level off. Collectively, this evidence implies that professional development should involve repeated practice in realistic settings in order to overwrite and upgrade existing habits.

**Corresponding author. Email address: s.sims@ucl.ac.uk*

Declarations of interest: none

1. Teacher effectiveness

Teachers' ability to raise pupil test scores follows a clear pattern over the career: increasing rapidly during the first three to five years, before growing more slowly in subsequent years (Papay & Kraft, 2015). Individual teacher growth curves, however, vary widely in terms of their gradient and the timing and level at which they flatten off (Kraft & Papay, 2014; Ost, 2014). As a result, teachers vary markedly in their ability to raise pupil achievement (Jackson, Rockoff, & Staiger, 2014). Indeed, a one standard deviation increase in the quality of teachers that a pupil is exposed to during high school is associated with a 0.3 standard deviation increase in test scores (Lee, 2018). Moreover, experimental research suggests that this reflects a causal relationship (Kane, McCaffrey, Miller, & Staiger, 2013). Understanding why teacher effectiveness begins to flatten off - and how it can be sustained - is therefore important for improving education.

A plausible explanation for the stabilisation of teacher effectiveness is the ossification of teaching practice. Indeed, a range of evidence suggests that classroom practice becomes resistant to change. For example, longitudinal studies have found that intensive teacher professional development programmes can fail to bring about changes in teacher practice, even in cases where teacher knowledge of the targeted new practices has increased (Copur-Gencturk & Papakonstantinou, 2016; Gess-Newsome et al., 2019). Certain pedagogical techniques which appear to be widely accepted within the teaching profession also appear to have had little impact on teachers' actions in the classroom (William, 2007). Moreover, studies using principal ratings of specific teacher skills based on classroom observations, show that these scores also begins to stabilise after around five years, suggesting practice stabilises around the same time as teaching effectiveness (Kraft, Papay, Chi, & Kraft, 2019).

The primary purpose of this article is to review and synthesise a recent body of evidence suggesting that habit formation is one important explanation for this phenomenon. While there has been little direct, rigorous research on teacher habits, we argue that there is already striking evidence for the prevalence of habitual behaviour among teachers. This appears not to have been recognised, however, because it requires combining theoretical and methodological insights about how to operationalise habit, drawn from neuroscience, with empirical findings about the way in which teachers' skills develop over time, drawn from the economics of education literature. We synthesise these findings, along with others from the psychology literature and new evidence on the automaticity of teaching practice drawn from a large survey of in-service teachers. This converging evidence provides strong reason to believe that habit formation plays an important part in limiting growth in teacher effectiveness.

One of the defining characteristics of habits is that they are insensitive to negative feedback, meaning that sub-optimal teaching practice can come to be locked in, even when they are known to be sub-optimal (Ost, 2014; Seger & Spiering, 2011). The primary motivation for this research is to understand how this can be avoided. Our hope is that, by providing an explanation for why growth in teacher effectiveness slows across the career, we can also shed light on how growth in teacher effectiveness can be maintained at higher levels, for longer. In particular, the article considers the implications of habitual practice for the design of teacher professional development, examining if, and how, habits can be replaced or upgraded.

Although we are primarily interested in teacher effectiveness, our focus on habitual practice also has relevance to issues of teacher professionalism. Schon (1983) argued that all professionals are granted autonomy to use their specialist knowledge to solve problems on behalf of society. In doing so, they must learn through a process of reflection on, and then adaptation of, their practice. Educationalists have since taken up this argument to emphasise that agency - teachers ability to shape their own work - is an important part of what makes them professionals (Biesta, Priestly, & Robinson, 2015; Heck, Grimmet, & Willis, 2019). The evidence we present on the importance of habits in constraining the potential for teachers to reflectively adapt their practice is therefore also relevant to broader debates on the nature of agency and professionalism in teaching.

The article is structured as follows. In section 2 we define habits and relate the characteristic features of habitual behaviour to observations of teaching practice in the education literature. In section 3, we

review the psychology literature on the conditions under which habits are most likely to be formed and compare these to known features of teaching. In section 4, we discuss how neuroscientists and psychologists have operationalised habits and review dovetailing empirical evidence drawn from the economics of education literature. In section 5, we present new evidence using self-report measures of habitual behaviour collected from a large cross-sectional sample of teachers in the UK. Following that, Section 6 draws out the implications for educational practice and section 7 concludes by discussing limitations of the existing literature and directions for future research.

2. Characteristics and neural basis of habitual behaviour

In this section, we will set out the underpinning neural basis for (non)habitual actions and explain how and why this gives habits their distinctive features. Specifically, habitual behaviours are characterised by ordered, structured action sequences that are automatically elicited by environmental cues and which are insensitive to goals or rewards.

2.1 Neural basis of habits

Importantly, goal-directed and habitual behaviours are associated with separate anatomical circuits. In humans, goal-directed behaviour is supported by a circuit connecting the prefrontal cortex (PFC) with two subcortical regions of the striatal basal ganglia: the caudate nucleus and the anterior putamen (Burton, Nakamura, & Roesch, 2015). In contrast, habitual behaviours rely on a loop connecting cortical regions in the sensorimotor and motor cortex with two further (though anatomically distinct) areas of the putamen, the medial and posterior putamen (Yin & Knowlton, 2006). The shift from goal-directed to habitual behaviour can therefore be observed through associated changes in neural activity between these loops. For example, extended practice of sequences of movements (conditions highly likely to produce habitual responding) can lead to decreased activation in goal-directed control circuits and increased activity in areas of the habit-related circuits such as the putamen (Lehericy et al., 2005; Steele & Penhune, 2010; Tricomi et al., 2009).

Causal inferences about the separate roles of these circuits can be made through animal learning studies, therefore a brief summary of animal evidence is important to illustrate the neural basis of such behaviour even though we naturally do not equate teaching performance with animal learning. For example, researchers have examined the effect of deactivating the rat dorsolateral striatum (DLS; an area homologous to the primate and human putamen, considered the “ground zero” for habit formation; Amaya & Smith, 2018) in paradigms where rats are trained to press a lever for a reward of sucrose, in order to produce habitual lever-pressing behaviour. Lesioning the DLS prior to any learning prevents the emergence of habitual behaviours and leads to rats rapidly reducing their lever pressing when their sucrose reward is devalued (Yin & Knowlton, 2006). In healthy rats, by contrast, once habitual behaviour has been established, devaluing the outcome does not affect the production of the behaviour (Ostlund et al., 2009). Further, if the DLS is deactivated after the lever-pressing habit has already been acquired, then behaviour rapidly reverts to being goal-directed, in the sense of being more sensitive to outcomes (Yin, Knowlton, & Balleine, 2006).

Thus, the development of habitual behaviour in both humans and animals has a neural basis, characterised by distinct patterns of activity in separate anatomical circuits. The patterns of activity within these circuits are also responsible for producing the characteristic features of habits, namely automaticity and goal insensitivity.

2.2 Habits are characterised by automaticity

Automaticity has been argued to be the ‘active ingredient’ of habit (Gardner, 2012). Habits are automatic in that they are activated in memory by external cues and then performed, without requiring executive control (Evans & Stanovich, 2013; Wood, Labreque, Lin, & Rüniger, 2014) and “without uncertainty or hesitation” (Saling & Phillips, 2007). Automaticity therefore refers to the relationship between environmental cues (outside the individual) and the production of a behaviour; in contrast to a process in which a behaviour is chosen to achieve a goal internal to the individual. Crucially, ‘a behaviour’ does not necessarily refer to a single action. When sequences of behaviours are performed repeatedly, eventually they may come to be produced too rapidly for each element in the sequence to

be cued by feedback from the preceding element (Lashley, 1951; Robbins & Costa, 2017). Central control of the sequence as a whole is therefore required, and as a result, single actions are combined into sequences which are stored as a single unit (Lungu et al., 2014; Steele & Penhune, 2010).

Habit learning mechanisms appear to be conserved across mammalian species, operating as strongly and consistently (and with the same limited input from higher cortical functions) in humans as in other mammals (Bayley, Frascino, & Squire, 2005). As a result animal studies, which allow for more invasive measurement and experimental manipulation than do studies in humans, provide one important form of evidence showing that habitual behaviours are characterised by automatically cued neural responses. Activity in the rat dorsolateral striatum (DLS; an area homologous to the primate and human putamen - an area essential for habit formation; see 'Neural basis of habits', section 2.3 below) displays automaticity in responding to a cue with an invariant sequence of activity patterns (Thorn, Atallah, Howe, & Graybiel, 2010). For example, the DLS develops an 'action-bracketing' activity pattern which chunks individual actions into a sequence (Smith & Graybiel, 2013). Thus, when a rat is presented with a particular habitual cue (such as a series of levers to press) after habits have been locked in by neural activity patterns, DLS neurons cease responding to each movement occurrence individually (Carelli, Wolske, & West, 1997), and instead signal whole action sequences automatically, leading to habitual sequences of behaviour (Ostlund, Winterbauer, & Balleine, 2009). The strength of the bracketing activity has been shown to correlate with faster and less deliberative (i.e. more automatic) behaviour patterns (Smith & Graybiel, 2013; Thorn et al., 2010).

Automaticity and habitual behaviour have also been linked in studies of humans. For example, research has found that if runners with strong running habits were subliminally cued with the contexts in which they typically run, they were faster to recognise the words 'running' and 'jogging'. Thus, stimulus-response associations could be retrieved faster in those with stronger habits (Neal, Wood, Labrecque, & Lally, 2012). Another study investigating the automatic cuing of habits by context found that participants with established habits of eating popcorn when at the cinema consumed more popcorn than others, even when the popcorn was stale and therefore unpleasant to eat (Neal, Wood, Wu, & Kurlander, 2011). The effect was reduced, however, by contextual changes which weakened the cuing of the behaviour, such as watching videos in a campus meeting room or eating with a non-dominant hand. Finally, repetition of an action (such as exercise, or healthy eating) in response to a stimulus has been shown to be associated with an increase in automaticity, as measured by a self-report questionnaire (Lally, Van Jaarsveld, Potts, & Wardle, 2010; Verplanken & Orbell, 2003).

Whilst little rigorous research has previously investigated automaticity in teaching, observational studies have documented experienced teachers performing complex classroom routines with "little explicit deliberation or forethought" (Vasquez-Levy, 1998) and without conscious interruption or modification (Leinhardt & Greeno, 1986; Wheatley & Wegner, 2001). These routines have been termed 'patterned action sequences' by researchers (Bohannon, 1995).

2.3 Habits are characterised by insensitivity

The automaticity of habitual behaviour leads to habitual actions being performed in a manner that is not goal-dependent and is therefore insensitive (at least initially) to changes in goals or the value of the response outcome. As such, habitual behaviours may have both positive and negative consequences. On the one hand, habits are essential to the efficient control of behaviour, because allowing regular actions to be performed automatically frees up cognitive resources for additional processing and facilitates multitasking (Taatgen, Huss, Dickison, & Anderson, 2008). In the classroom, where feeling overwhelmed by the levels of simultaneous activity is a common experience for novice teachers (Carré, 1993; Corcoran, 1981; Kagan, 1992; Olson & Osborne, 1991; Veenman, 1984; Wideen, Mayer-Smith, & Moon, 1998), developing effective habits will be essential to avoiding cognitive overload (Feldon, 2007). On the other hand, habitual behaviours can sometimes be prompted automatically by contextual cues in situations where they are unwanted or counterproductive.

Animal studies have demonstrated the insensitive nature of habitual behaviours through establishing a 'devaluation of outcome effect'. This involves measuring the extent to which the devaluation of an

outcome (such as replacing a desirable reward with an unpleasant or noxious one, or an ‘extinction phase’ where rewards are withheld) leads to a change in the rate of production of the behaviour. The degree that a behaviour remains present even following changes in rewarding outcomes provides a behavioural measure of habit performance. Outcome-insensitivity has been demonstrated in animals in this using various paradigms. For example, following outcome devaluation procedures rats have been reported to continue to perform head-movements (Tang, Pawlak, Prokopenko, & West, 2007), follow routes in a maze (De Leonibus et al., 2011; Smith & Graybiel, 2016), track previously rewarded stimuli (such as coloured lights) (Morrison, Bamkole, & Nicola, 2015), and press levers (Coutureau & Killcross, 2003; Dickinson, 1985; Smith & Graybiel, 2013), despite the devaluation of the outcome.

Insensitivity to outcome devaluation has also been experimentally induced in humans through overtraining in a study in which participants learned to press keys in order to obtain snack food rewards (Tricomi, Balleine, & O’Doherty, 2009). Following training, outcome devaluation was induced through selectively satiating participants on one of the two available snack foods (i.e. consuming them until they were no longer desired), after which participants were once again tested on the trained paradigm. Participants in an extensive training group (3-days) consumed more of the devalued reward (and so were less sensitive to the devaluation) than a group trained for only one day (although c.f. one recent failure to replicate this specific overtraining effect; de Wit et al., 2018). Evidence for the goal insensitivity of habitual behaviour in humans can also be found in the social-psychology literature. Meta-analyses of studies looking at the relationship between intentions and behaviour have found that intentions are stronger predictors of actions that were performed less frequently and that habits are a strong (negative) moderator of the links between goals and behaviours (Gardner, De Bruijn, & Lally, 2011; Ouellette & Wood, 1998; Webb & Sheeran, 2006). Thus, habitual behaviours are frequently performed in spite of intentions or goals to behave differently.

While evidence from educational settings is again more scarce, teachers’ behaviours have also been shown to be insensitive to a change in goals. In one notable example, 36 experienced high school teachers were asked to teach two consecutive (observed) lessons to small groups of randomly-selected students (Doyle & Redwine, 1974). Following the first session, teachers received feedback on the extent to which their lesson had followed their own planned time allocations, but this feedback was manipulated to suggest either a high or low level of discrepancy with the original plans. The participants then submitted a revised plan of their intentions for the second session on the basis of this feedback, before their second lesson was observed. Results showed that, regardless of the degree of intent to change in the second session, the teachers did not measurably change their practice.

2.4 Summary

In summary, habits involve the automatic production of behaviour(s) in responses to cues in the environment, in a way that is insensitive to goals or rewards. Both the automaticity and insensitivity of habitual behaviours is underpinned by the separate neurological circuits which regulate them. Transferring regulation of an activity from the goal-directed, deliberative circuit frees up scarce resources in working memory to aid further problem solving (Seger & Spiering, 2011). However, shifting regulation of a behaviour to automatic, insensitive circuit can lock in sub-optimal behaviour (Wood & Runger, 2016). This has been described by educational psychologists as the “double-edged sword of automaticity” (Feldon, 2007).

3. Classroom contexts and the development of habitual behaviour

In this section, we review evidence on whether teachers’ working environment is conducive to the development of habit. A range of factors have been shown to moderate the production of habitual behaviours¹, some of which are very relevant when considering the potential impact of habits on teaching practice. Amongst these, habit performance has been shown to increase in conditions of frequent performance of the behaviour, under time or performance pressure and under conditions of stress. The following sections examine each of these factors in turn.

3.1 Frequent responding promotes habit formation

Although frequently performing a behaviour does not guarantee future automaticity (Gardner, 2012; Wheatley & Wegner, 2001), research has found a relationship between frequency of repetition and habit development (Snihotta & Pesseau, 2012). A clear example of this principle is the observation that habits are most likely to form for behaviours which are performed most frequently in our everyday lives, such as behaviours related to eating, drinking, purchasing, exercising or driving (Aldrich, Montgomery, & Wood, 2011; de Bruijn, Kremers, Singh, van den Putte, & van Mechelen, 2009; Ji & Wood, 2007; Ouellette & Wood, 1998; Webb & Sheeran, 2006). In general, the frequency of past behaviour is a far weaker predictor of future behaviour in domains that are encountered only annually or biannually than for those which are encountered on a daily or weekly basis (Ouellette & Wood, 1998). Self-reported habit strength also correlates strongly with the frequency of past behaviour (Verplanken & Orbell, 2003). Furthermore, when high frequency behaviours are performed in stable contexts, intentions are very poor predictors of behaviour (Danner, Aarts, & De Vries, 2008).

Many aspects of teaching are highly repetitive. For example, teachers in England teach around 950 lessons per year (National Union of Teachers, 2008), all of which (excluding double lessons) involve executing certain routines such as getting a class into the room and settled at their desks. Many teachers will go on to re-teach these same lessons in subsequent academic years. Within lessons, teachers also repeat certain actions many times. For example, systematic observations have recorded teachers asking between 49 and 89 questions in a one hour lesson, which amounts to tens of thousands of questions per year (Hargreaves, 1984; Resnick, 1972). Teachers also expend considerable effort keeping order during the lesson, with studies recording an average of 16 instances of observable behavioural reprimands per 45 minute lesson - equivalent to around 20,000 instances per year (Wheldall, Houghton, & Merrett, 1989). Teaching therefore involves repetitive responding to recurring environmental cues.

3.2 Time or performance pressures promote habitual behaviour

In laboratory studies, participants placed under time pressure have been found to perform tasks by relying less on deliberative processes and more on habit-type processes (Toth, 1996; Yonelinas & Jacoby, 1994; 1995). For example, in cued-recall studies, where participants are required to respond to cues to complete a task (following an initial learning period), time pressure leads to a reduced reliance on conscious recall, in favour of automatic processes (Yonelinas & Jacoby, 2012). Similar effects can be produced through inducing performance pressure, for example by taxing participants with a secondary assignment on top of their main task, which leads to increased habitual actions, even when this is detrimental to overall performance (Ruh, Cooper, & Mareschal, 2010).

Teachers working in many different countries and school systems face strong accountability pressures through mechanisms of parental school choice (Mattei, 2012) and from official school inspectorates (Ehren et al., 2015). These pressures are perhaps most extreme in the US and UK, where school leaders are regularly removed from their post based on the schools exam results (Eyles & Machin, 2019; Heissel & Ladd, 2018). This puts school leaders under pressure, which is then communicated downwards to the teaching staff (Leithwood, Steinbach, & Jantzi, 2002; Perryman, Ball, Maguire, & Braun, 2011; Wise, 2001). These accountability pressures have been linked with increases in workload as teachers attempt to respond to the demands of management (Perryman & Calvert, 2019), and international teacher surveys show long hours are common in OECD countries (Jerrim & Sims, 2019). Teaching is thus subject to strong accountability and time pressures.

3.3 Stress promotes habitual behaviour

When time, accountability or other pressures increase to the point that an individual feels they may not be able to achieve their goals, this causes stress (Schuler, 1980) which is known to have wide ranging effects on human learning and memory systems (Roosendaal, McEwen, & Chattarji, 2009; Schwabe, Wolf, & Oitzl, 2010). In particular, stress can limit the capacity for deliberative thought, resulting in a greater likelihood of automatic and habitual responding to a stimulus (Schwabe et al., 2007). The mediating mechanism here is likely to be overlapping with that linking time and

accountability pressures and habitual responding, but the primary causes of these in teaching are distinct (Berryhill, Linney, & Fromewick, 2009).

In one experimental demonstration of this effect, participants underwent a learning phase where they were asked to choose between different symbols presented on a computer screen, which related to food outcomes: chocolate milk, orange juice, peppermint tea or water (Schwabe & Wolf, 2010). The food outcomes had different probabilities of presentation, depending on the trial type (meaning that participants had to learn the connection between the symbols presented and the likelihood of receiving particular reward types). After the learning session, either the chocolate or the orange rewards were devalued (through allowing participants to satiate on either chocolate pudding or oranges). Before participants were re-exposed to the original task, however, some participants underwent a cold-pressor task (immersing their hand in ice water, under the supervision of an unfriendly experimenter) designed to heighten their stress levels. Results showed that stressed participants were insensitive to the outcome devaluation compared to controls, that is, stress led to an increase in habitual behaviour. Cortisol responses to stress also correlated significantly with habitual performance. Interestingly, a similar procedure which introduced stress *before* any learning occurred also found that this promoted habitual responding (Schwabe & Wolf, 2009). Stress therefore seems able to facilitate both the formation and production of habitual behaviours.

Teaching is widely recognised to be a stressful job (Cooper & Travers, 2012; Johnson et al., 2005). Indeed, survey research repeatedly finds it to be among the most stressful of any major occupational group (Johnson et al., 2005; Smith, Brice, Collins, Matthews, & McNamara, 2000). There is also evidence beyond self-report data, with studies showing that early-career teachers show markedly elevated cortisol levels after conducting a lesson (Wolfram, Bellingrath, Feuerhahn, & Kudielka, 2013). Thus, teachers are subject to relatively high levels of stress, which is likely to impede deliberate reflection and adaptation of the sort thought to be crucial for their development as professionals (Biesta, Priestly, & Robinson, 2015; Heck, Grimmet, & Willis, 2019).

4. Evidence from performance asymptotes

4.1 Operationalising habit using performance asymptotes

As discussed, habitual behaviours are more likely to emerge in situations where stimulus-response-outcome contingencies are stable, in other words in situations where a consistent response is produced to the same stimulus, resulting in a consistent outcome. Stability of behaviour naturally also leads to stability of performance, so the emergence of habitual behaviours can be tracked through the slowing of learning to the point where performance approaches asymptote (Dickinson, 1985). Indeed, the only other explanation would be that practice continued to change but somehow produced the exact same level of performance. Frequently therefore, in neuroscience and cognitive psychology research, the habit formation process is operationalised as the point at which the learning curve (measuring accuracy or completion time for a task) levels off, or approaches asymptote (Brasted & Wise, 2004; Brovelli, Nazarian, Meunier, & Boussaoud, 2011; Daw, Niv, & Dayan, 2005; de Wit et al., 2018; Schwabe & Wolf, 2010; Zwosta, Ruge, Goschke, & Wolfensteller, 2018).

Crucially, for our argument, activity in brain regions associated with goal-directed behaviour, such as the caudate (see section 2.3 *Neural basis of habits*), has been found to closely track position on a learning curve. For example, Williams and Eskandar (2006) recorded neural activity from monkey caudate whilst they were trained to associate images with joystick movements in a specific direction. Caudate activity was high during the upward sloping part of the learning curve but then began to fall just before performance approached asymptote (Figure 1, Panel B; see also Brasted & Wise, 2004; Brovelli et al., 2011). Similarly, in an experiment using human subjects, in which participants learned to categorise visual stimuli as either 'natural' or 'artificial' using a key press, improved performance was shown to be associated with increased functional connectivity between posterior putamen and premotor cortex, suggesting increasing strength of habit formation (Zwosta et al., 2018).

It is important to note here that the correlation between activity in habit circuits and approaching performance asymptote (and so habitual behaviour) is not evidence that the activity in these circuits

causes such habits to emerge. Indeed, there is evidence to suggest that the ‘switch’ from goal-directed to habitual behaviour is also a function of the remaining involvement (or not) in goal-directed circuits, as much as increases in the activity of habit networks (Smith & Graybiel, 2013; Zwosta et al., 2018). Equally, while the ‘habit’ and ‘goal-directed’ circuits are necessary for the production of their associated behaviour, they are not entirely sufficient, relying both on each other and on a range of other brain regions (Steele & Penhune, 2010), as with any complex behaviour. What is important to establish for present purposes, however, is that approaching performance asymptote is often closely associated with the emergence of habitual behaviour, and that this process is associated with distinct and measurable patterns of neural activity.

4.2 Performance asymptotes in the teaching literature

Laboratory neuroscience studies of habit generally employ simple tasks in which performance is easy to measure, such as completion of a maze (Smith & Graybiel, 2013) or lever pressing (Yin et al., 2006). In the field, perhaps the earliest study of habitual behaviour - *Studies on the Telegraphic Language: The Acquisition of a Hierarchy of Habits* was conducted on Morse code operator trainees. Performance in this setting can be quantified as the number of characters correctly transmitted or received per minute and the results, published in 1899, showed an asymptotic learning curve (William & Harter, 1899). Similar studies have since been conducted in a range of settings, for example with air traffic controllers training using simulators (Joyanovic & Nyorko, 1995). In this setting, performance can be measured as the number of planes safely landed per period of time and the results, again, showed a similar asymptotic learning curve.

More recently, the advent of annual pupil testing in the US has allowed researchers to estimate the contribution of individual teachers to pupil progress. Although we acknowledge that raising test scores is far from the only important role played by teachers (DeAngelis, 2019; Harris & Sass, 2014), using test scores as a measure of performance has the benefit of providing a precise quantification of effectiveness. Moreover, these annual value added measures (VAM) have been shown to correlate well with teacher’s lifetime VAM (Staiger & Kane, 2015), measures derived from more complex cognitive tasks (Kraft, 2019), pupils earnings in later life (Chetty, Friedman, & Rockoff, 2014), and VAM measures derived from random assignment of pupils to teachers (Kane et al., 2013). Modelling growth in teacher VAM over time is not straightforward because it suffers from the age-period-cohort problem, which requires that assumptions be invoked (Papay & Kraft, 2015). Different approaches employ different assumptions and estimate curves with slightly different shapes as a result. Within the degree of precision afforded by the data however, all methods show teacher effectiveness rising quickly during the early years of the career and growing at a slower and slower rate thereafter – that is to say, an asymptotic curve (Kraft & Papay, 2014; Ost, 2014; Papay & Kraft, 2015).

Taken together with the evidence from Section 2 and 3 – particularly the research showing that tasks are handed off to the habit circuit at the point the curve approaches asymptote – we interpret this as further converging evidence that habits limit the increase in teacher effectiveness. Two important caveats are in order at this point. First, unlike lever pressing or maze completion, teaching skill is a composite, made up of many constituent subskills. The levelling off of the learning curve among teachers therefore represents an aggregate of many skill-specific learning curves, each of which are likely to approach asymptote at different points. Second, we would like to emphasise again that raising test scores is an incomplete measure effectiveness, that misses other aspects of what it is that teachers do.

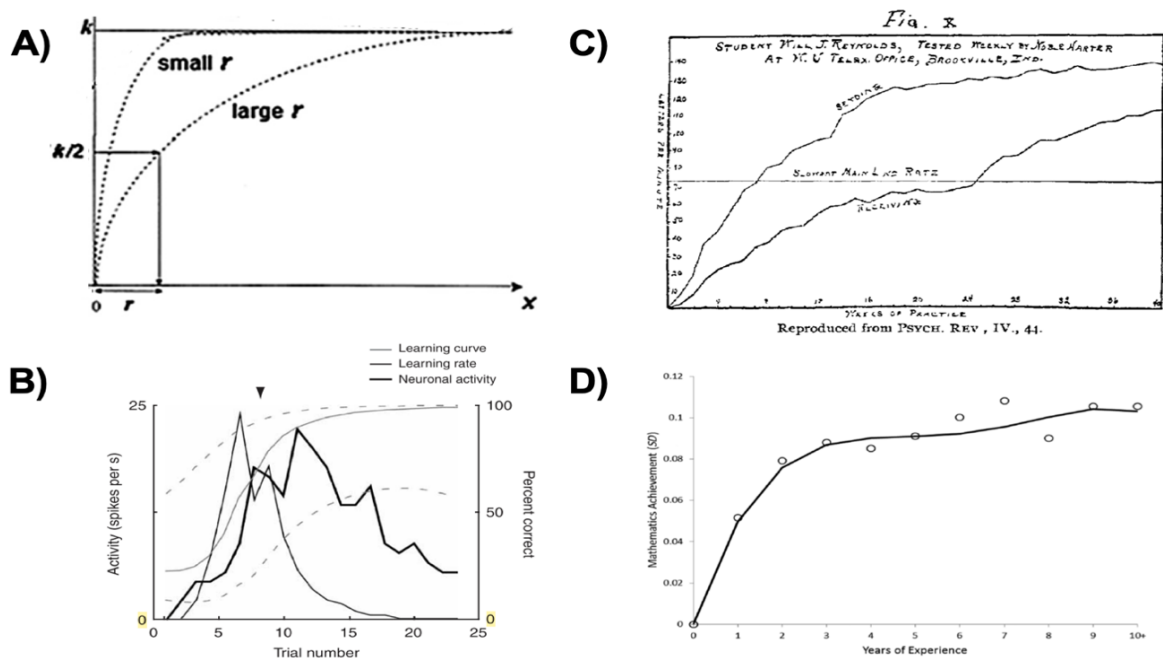


Fig. 1. Learning curves depicting performance asymptotes across a variety of explanatory levels. A) mathematical depiction of a learning curve in which r represents the learning rate (with a smaller r describing faster learning), k represents the asymptotic limit of growth (so a larger k implies higher performance at asymptote) and the equation for performance is $= k \left(\frac{\exp}{\exp + r} \right)$ (Anzanello & Fogliatto, 2011). B) Performance accuracy and mean neuronal activity in the caudate during a learning task (Williams & Eskandar, 2006). Grey line, monkeys' behavioural performance (learning curve); dashed lines, upper and lower confidence bounds (99%) estimated from the monkeys' performance. Black line, learning rate. Thick black line, average firing rate for the cell during the period. Caudate activity closely correlates with the learning curve until it reaches asymptote, then falls away as performance stabilises. C) Morse code operative performance over time (William & Harter, 1899). Letters per minute sent by weeks of practice. Learning curves shown for sending (upper line) and receiving (lower line). D) Learning curve for an average teacher in terms of effect on pupil exam results by years of experience (Kraft & Papay, 2014).

5. Direct evidence of automaticity among teachers

So far we have argued that many of the characteristic features of habitual behaviours can be observed in teachers and that a number of factors known to facilitate the formation and reproduction of habits are present in schools and classrooms. However, research has not previously attempted to directly measure habitual behaviour in teaching. In this section, we provide new evidence on this point using data from a large cross-sectional survey of teachers in the UK.

5.1 Self-report measures of habitual behaviour

Verplanken and Orbell developed and validated the 12-item Self-Report Habit Index (SRHI) as a questionnaire-based alternative to behavioural frequency measures of habit (Gardner & Tang, 2014; Verplanken & Orbell, 2003). The SRHI has since been used in wide range of studies, particularly in the behavioural medicine literature, where it has been used to study nutrition, exercise and use of active transport modes (Gardner et al., 2011). More recently, a short four-item automaticity subscale of the index (the Self-Report Behavioural Automaticity Index, or SRBAI) has been developed and validated (Gardner, Abraham, Lally, & de Bruijn, 2012). This has been found to be a reliable and sensitive measure of habitual behaviour and shown to predict two key effects of habit on behaviour: a correlation between habit strength and behaviour, and a moderating effect of habit on the relationship

between intention and behaviour (Gardner et al., 2012). Respondents using the SRBAI are asked to respond to four items on a ‘strongly agree’ to ‘strongly disagree’ scale, in relation to a specific behaviour: ‘*is something I do automatically*’; ‘*is something I do without having to consciously remember*’; ‘*is something I do without thinking*’ and ‘*is something I start doing before I realize I’m doing it*’. The result is a measure of habit strength which retains the SRHI’s predictive validity for behaviour, whilst allowing for the more parsimonious investigation of habitual behaviours.

We adapted the SRBAI to investigate automaticity in teaching. In order to get a clean estimate of the relationship between experience and automaticity, we wanted to pick behaviours which teachers were unlikely to have any experience with before entering the profession. Questioning techniques, for example, might be something that could be learned while pupils were still at school, perhaps through debating societies or through pupils observing their own teachers. We also wanted to choose target behaviours which could be described succinctly and with minimal ambiguity. We therefore chose two target behaviours/sequences: getting students into the room at the start of a lesson “Getting students settled at the start of a lesson...” and managing pupil behaviour “Managing low-level disruption...”. Based on the literature reviewed in Sections 2-4, we hypothesised that teachers would show increasing automaticity on these skills during their early careers, levelling off as the behaviours came to be regulated by the neural circuits involved in habitual behaviour.

5.2 Data collection

We administered the questionnaire to teachers via an online, multiple-choice, UK-based teacher survey app called Teacher Tapp. Teacher Tapp only asks users three questions per day, so the eight items were administered over multiple days in April 2019. Responses were collected using a 7-point Likert scale. Our measures of teacher experience were also derived from Teacher Tapp and are therefore based on multiple choice, categorical responses to a question about the number of years since qualification. Although the data represents a convenience sample, using Teacher Tapp allowed us to collect data from an unusually large sample of teachers: yielding 1,337 complete responses to the question on managing low-level disruption and 1,244 complete responses to the question on getting students settled. The responses to the four items for each construct were converted to a single score using confirmatory factor analysis, with alphas of 0.85 and 0.9 respectively, suggesting high internal consistency for the scales. The factor scores were normalised to have a mean of 0 and a standard deviation of one.

5.3 Results

Figure 2 shows the mean factor score for various experience bins, along with a 95% confidence interval. The y axis is scaled to reflect the full range of the factor scores. Both panels of the chart show a similar picture in which teachers with very low level of experience report automaticity around 0.75 standard deviations below the mean, which rises to be close to the average after between three and five years on the job. At this point, automaticity begins to level off, remaining stable in the 5-10 and 10-20 years of experience categories. These findings are consistent with our hypothesis that automaticity would rise initially and then level off. The timing of the curve approaching asymptote in automaticity is also approximately consistent with the timing of performance reaching asymptote seen in Section 4.

These findings have a number of limitations, not least that we are only measuring the automaticity component of habits and are focusing on just two target behaviours – both of which are related to classroom management. Nevertheless, we interpret these results as further converging evidence that the gradual development of habitual behaviours serves to limit increases in teacher effectiveness over time.

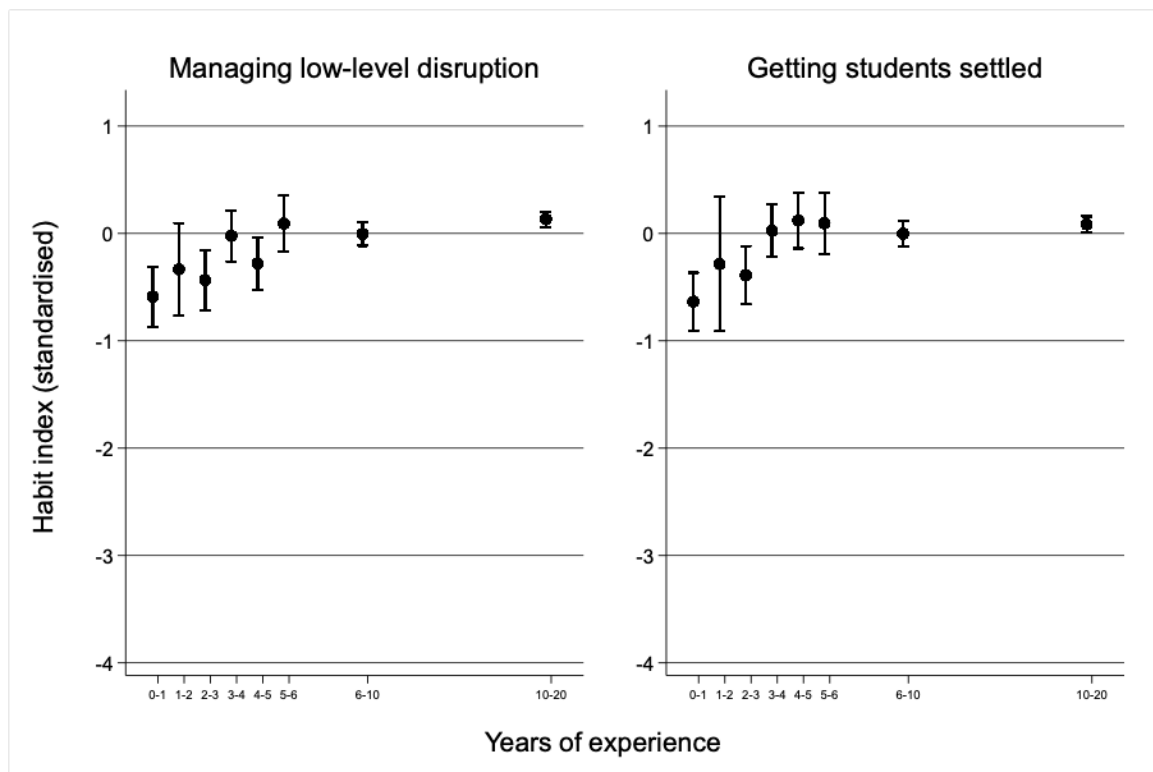


Fig. 2. Mean of the Self-Report Behavioural Automaticity Index factor score plotted against categories of teaching experience. Dots represent the mean and whiskers represent the 95% confidence interval. Left hand side N=1,337; right hand side N=1,224.

6. Discussion of implications for practice

It is important to reemphasise at this point that habitual control of behaviour in classrooms is not necessarily a bad thing. Developing automatic routines is an essential step for novice teachers if they are to begin to overcome the high levels of cognitive load that can be imposed by working in the classroom (Feldon, 2007). However, whilst a degree of automaticity can be beneficial to a teacher, it is clear that rigidity in behaviour can be maladaptive. As discussed above, the immediate and automatic triggering of habits by associated contexts means that habitual behaviours may override more deliberative intentions. When such intentions conflict with the established habit, behaviour is likely to proceed in line with the habit, rather than the intention (Webb & Sheeran, 2006). The stronger the habit, the less intention influences performance (Verplanken, Aarts, Van Knippenberg, & Moonen, 1998). This may lead to an inability to adapt practice in the light of new evidence on effective practice, new curricula or different types of students – thus locking in certain suboptimal ways of working.

Clearly, these findings have implications for teacher development. In particular, they suggest that professional development programs that target increased teacher knowledge of certain pedagogical techniques are unlikely to be sufficient for improved practice. Unless environmental changes which serve to weaken the contextual cues for pre-existing habitual behaviour are also enacted in parallel, increasing teachers knowledge is unlikely to disrupt established cue-response associations (Gardner, 2012; Webb & Sheeran, 2006). Unfortunately, this approach seems to be common in education - data from international teacher surveys shows that for two thirds of teachers (66.6%) most professional development only involves listening to an instructor provide information (OECD, 2013: Table 4.19). Similarly, alerting teachers to the difficulty of changing practice is unlikely to help, since habitual responses are likely to be triggered and possibly even executed before a teacher has become consciously aware of this fact. Because habits are regulated by a circuit in the brain characterised by

fast, automatic responses, they will likely always be *one step ahead* of conscious processes regulated by the slow, deliberative circuit.

Given the difficulties involved in trying to overcome habit by conscious means, it is likely more promising to focus instead on replacing one habitual response with another. This involves re-associating the same cues with new responses thus short-circuiting pre-existing cue-response links. Replacing habits can be achieved in the same way that the original habits were established: through repetition of the desired behaviour in the presence of the relevant environmental cues (Ouellette & Wood, 1998; Webb & Sheeran, 2006). With respect to teaching, this would require the repeated use of a new technique or approach in a realistic (or near-realistic) classroom environment.

One type of professional development which incorporates these features is instructional coaching, which involves experts working with teachers to develop specific skills in a sustained observation-feedback-practice cycle (Knight, 2007; Kraft & Blazar, 2017; Kraft, Blazar, & Hogan, 2016). Coaching often involves rehearsing new techniques in a controlled environment, then implementing them in a real lesson (often captured on video), then receiving feedback from an observer and then reimplementing the technique in a subsequent lesson. Interestingly, instructional coaching has been shown in replicated randomised controlled trials (Allen, Hafen, Gregory, Mikami, & Pianta, 2015; Allen, Pianta, Gregory, Mikami, & Lun, 2011) and in meta-analysis of causal studies (Kraft et al., 2016) to bring about changes in teaching practice and improvements in pupil test scores. Of course, there are many other types of professional development which could incorporate repeated practice in the classroom environment, but we interpret the literature on instructional coaching as further converging evidence that habits limit teacher effectiveness - unless they are replaced or overwritten with new habits.

7. Conclusion

We have argued that growth in teacher effectiveness declines over time in large part because teachers' practice becomes habitual. That is, practice comes to be triggered automatically by external cues in a way that is insensitive to goals or payoffs. This argument has been developed through a review of converging evidence from neuroscience, social science and education literature. We have synthesised evidence suggesting that teaching is highly conducive to habit formation and that teachers display characteristic features of habitual behaviour. In addition, we have highlighted how findings from neuroscience – that a handover of behaviour regulation from a goal directed circuit to a habit circuit occurs at the point at which performance approaches asymptote – directly supports the interpretation that teacher learning curves level off because of habit formation. Further, we have presented new data derived from self-report measures among practising teachers showing that automaticity of behaviour mirrors these performance asymptotes. In sum, we believe that this converging evidence provides strong reason to believe that habit formation is an important factor limiting teacher effectiveness.

Having said this, we also acknowledge a number of limitations of the existing evidence. Foremost amongst these is a lack of direct evidence regarding habit formation and production in teachers. In particular, although the data presented in Section 5 provides direct evidence on automaticity of behaviour among teachers, the SRBAI does not capture the other important characteristic of habits: insensitivity. Future research should explore insensitivity among teachers, either by using alternative self-report measures or through experimental manipulations, ideally using teachers conducting teaching-related tasks in classroom environments. A second important limitation of the original data presented in Section 5 is that it only refers to two specific teaching behaviours/sequences, both of which relate to behaviour management. Investigating automaticity across a wider range of teaching behaviours would help clarify whether our findings hold more generally. Understanding which behaviours are most prone to automaticity and at which point in teachers' careers this occurs could have important implications for the design of teacher professional development.

In spite of these limitations, the evidence synthesised here has a number of implications for policy and practice. Foremost amongst these is that habits may come to lock in certain suboptimal practices, which have become automatic. We have already discussed the importance of designing professional development to include repeated practice in realistic settings in order to overwrite and upgrade such

habitual practices. Survey research suggests that much existing professional development does not currently do this, implying significant potential for improvement. In addition, the article highlights important tensions between holding teachers and schools accountable for poor performance, and the potential for them to improve ingrained patterns of practice while under pressure from the accountability system. It is possible that the stressful nature of very high stakes accountability systems serves to constrain rather than induce changes in teaching practice. Again, further research is required to understand the trade-offs involved here. Relatedly, school leaders should be careful not to overload teachers with administrative or other non-teaching tasks in order to ensure they do not experience excessive time pressure. Taken together, these implications have the potential to sustain the upward trajectory of teachers' learning curves longer into their careers, thus benefiting pupils.

References

- Aldrich, J. H., Montgomery, J. M., & Wood, W. (2011). Turnout as a Habit. *Political Behavior*, 33(4), 535–563. <https://doi.org/10.1007/s11109-010-9148-3>
- Allen, J. P., Hafen, C. A., Gregory, A. C., Mikami, A. Y., & Pianta, R. (2015). Enhancing Secondary School Instruction and Student Achievement: Replication and Extension of the My Teaching Partner-Secondary Intervention. *Journal of Research on Educational Effectiveness*, 8(4), 475–489. <https://doi.org/10.1080/19345747.2015.1017680>
- Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science*, 333(6045), 1034–1037. <https://doi.org/10.1126/science.1207998>
- Allen, R., Benhanda, A., Jerrim, J., & Sims, S. (2019). Teacher working hours: evidence from four datasets. *Under Review*.
- Amaya, K. A., & Smith, K. S. (2018). Neurobiology of habit formation. *Current Opinion in Behavioral Sciences*, 20, 145–152. <https://doi.org/10.1016/j.cobeha.2018.01.003>
- Anzanello, M. J., & Fogliatto, F. S. (2011). Learning curve models and applications: Literature review and research directions. *International Journal of Industrial Ergonomics*, 41(5), 573–583. <https://doi.org/10.1016/j.ergon.2011.05.001>
- Bailey, P. J., Frascino, J. C., & Squire, L. R. (2005). Robust habit learning in the absence of awareness and independent of the medial temporal lobe. *Nature*, 436(7050), 550–553. <https://doi.org/10.1038/nature03857>
- Berryhill, J., Linney, J. A., & Fromewick, J. (2009). The Effects of Education Accountability on Teachers: Are Policies Too Stress-Provoking for Their Own Good? *International Journal of Education Policy and Leadership*, 4(5), 1–1. <https://doi.org/10.22230/ijep.2009v4n5a99>
- Biesta, G., Priestley, M., & Robinson, S. (2015). The role of beliefs in teacher agency. *Teachers and Teaching*, 21(6), 624–640.
- Bohannon, P. (1995). *How Culture Works*. New York: Free Press.
- Brasted, P. J., & Wise, S. P. (2004). Comparison of learning-related neuronal activity in the dorsal premotor cortex and striatum. *European Journal of Neuroscience*, 19(3), 721–740. <https://doi.org/10.1111/j.0953-816X.2003.03181.x>
- Brovelli, A., Nazarian, B., Meunier, M., & Boussaoud, D. (2011). Differential roles of caudate nucleus and putamen during instrumental learning. *NeuroImage*, 57(4), 1580–1590. <https://doi.org/10.1016/j.neuroimage.2011.05.059>
- Burton, A. C., Nakamura, K., & Roesch, M. R. (2015, January 1). From ventral-medial to dorsal-lateral striatum: Neural correlates of reward-guided decision-making. *Neurobiology of Learning and Memory*. Academic Press. <https://doi.org/10.1016/j.nlm.2014.05.003>
- Carelli, R. M., Wolske, M., & West, M. O. (1997). Loss of lever press-related firing of rat striatal forelimb neurons after repeated sessions in a lever pressing task. *Journal of Neuroscience*, 17(5), 1804–1814. <https://doi.org/10.1523/jneurosci.17-05-01804.1997>
- Carré, C. (1993). The first year of teaching. In N. Bennett & C. Carré (Eds.), *Learning to teach* (pp. 119–211). New York: Routledge.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher

- value-added and student outcomes in adulthood. *American Economic Review*, 104(9), 2633–2679. <https://doi.org/10.1257/aer.104.9.2633>
- Cooper, C., & Travers, C. (2012). *Teachers under pressure: Stress in the teaching profession*. Routledge.
- Copur-Gencturk, Y., & Papakonstantinou, A. (2016). Sustainable changes in teacher practices: a longitudinal analysis of the classroom practices of high school mathematics teachers. *Journal of Mathematics Teacher Education*, 19(6), 575–594. <https://doi.org/10.1007/s10857-015-9310-2>
- Corcoran, E. (1981). Transition Shock: The Beginning Teacher's Paradox. *Journal of Teacher Education*, 32(3), 19–23. Retrieved from <https://eric.ed.gov/?id=EJ249374>
- Coutureau, E., & Killcross, S. (2003). Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. *Behavioural Brain Research*, 146(1–2), 167–174. <https://doi.org/10.1016/j.bbr.2003.09.025>
- Danner, U. N., Aarts, H., & De Vries, N. K. (2008). Habit vs. intention in the prediction of future behaviour: The role of frequency, context stability and mental accessibility of past behaviour. *British Journal of Social Psychology*, 47(2), 245–265. <https://doi.org/10.1348/014466607X230876>
- Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8(12), 1704–1711. <https://doi.org/10.1038/nn1560>
- DeAngelis, C. A. (2019). Divergences between effects on test scores and effects on non-cognitive skills. *Educational Review*, 1-12.
- de Bruijn, G. J., Kremers, S. P. J., Singh, A., van den Putte, B., & van Mechelen, W. (2009). Adult Active Transportation. Adding Habit Strength to the Theory of Planned Behavior. *American Journal of Preventive Medicine*, 36(3), 189–194. <https://doi.org/10.1016/j.amepre.2008.10.019>
- De Leonibus, E., Costantini, V. J. A., Massaro, A., Mandolesi, G., Vanni, V., Luvisetto, S., ... Mele, A. (2011). Cognitive and neural determinants of response strategy in the dual-solution plus-maze task. *Learning and Memory*, 18(4), 241–244. <https://doi.org/10.1101/lm.207431>
- de Wit, S., Kindt, M., Knot, S., Verhoeven, A., Robbins, T., Gasull-Camos, J., ... Gillan, C. (2018). Shifting the balance between goals and habits: five failures in experimental habit induction. *Journal of Experimental Psychology-General*, 147(7), 1043–1065.
- Dickinson, A. (1985). The Development of Behavioural Autonomy. *Philosophical Transactions of the Royal Society of London . Series B*, 308(1135), 67–78. Retrieved from <https://www.jstor.org/stable/2396>
- Doyle, W., & Redwine, J. M. (1974). Effect of intent-action discrepancy and student performance feedback on teacher behavior change. *Journal of Educational Psychology*, 66(5), 750–755. <https://doi.org/10.1037/h0037476>
- Ehren, M. C. M., Gustafsson, J. E., Altrichter, H., Skedsmo, G., Kemethofer, D., & Huber, S. G. (2015). Comparing effects and side effects of different school inspection systems across Europe. *Comparative Education*, 51(3), 375–400. <https://doi.org/10.1080/03050068.2015.1045769>
- Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition. *Perspectives on Psychological Science*, 8(3), 223–241. <https://doi.org/10.1177/1745691612460685>
- Eyles, A., & Machin, S. (2019). The Introduction of Academy Schools to England's Education. *Journal of the European Economic Association*, 17(4), 1107–1146. <https://doi.org/10.1093/jeea/jvy021>
- Feldon, D. F. (2007). Cognitive Load and Classroom Teaching: The Double- Edged Sword of Automaticity. *Educational Psychologist*, 42(3), 123–137. <https://doi.org/http://dx.doi.org/10.1080/00461520701416173>
- Gardner, B. (2012). Habit as automaticity, not frequency. *The European Health Psychologist*, 14(2), 32–36. <https://doi.org/10.1037/e544772013-003>
- Gardner, B., Abraham, C., Lally, P., & de Bruijn, G.-J. (2012). Towards parsimony in habit measurement: Testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. *International Journal of Behavioral Nutrition and Physical Activity*, 9(102). <https://doi.org/10.1186/1479-5868-9-102>
- Gardner, B., De Bruijn, G. J., & Lally, P. (2011). A systematic review and meta-analysis of

- applications of the self-report habit index to nutrition and physical activity behaviours. *Annals of Behavioral Medicine*, 42(2), 174–187. <https://doi.org/10.1007/s12160-011-9282-0>
- Gardner, B., & Tang, V. (2014). Reflecting on non-reflective action: An exploratory think-aloud study of self-report habit measures. *British Journal of Health Psychology*, 19(2), 258–273. <https://doi.org/10.1111/bjhp.12060>
- Gess-Newsome, J., Taylor, J. A., Carlson, J., Gardner, A. L., Wilson, C. D., & Stuhlsatz, M. A. M. (2019). Teacher pedagogical content knowledge, practice, and student achievement. *International Journal of Science Education*, 41(7), 944–963. <https://doi.org/10.1080/09500693.2016.1265158>
- Hargreaves, D. H. (1984). Teachers' questions: Open, closed and half-open. *Educational Research*, 26(1), 46–51. <https://doi.org/10.1080/0013188840260108>
- Harris, D. N., & Sass, T. R. (2014). Skills, productivity and the evaluation of teacher performance. *Economics of Education Review*, 40, 183–204.
- Heck, D., Grimmer, H., & Willis, L. D. (2019). Teacher Educators Using Cogenerative Dialogue to Reclaim Professionalism. In *Professionalism and Teacher Education* (pp. 137–156). Springer, Singapore.
- Heissel, J. A., & Ladd, H. F. (2018). School turnaround in North Carolina: A regression discontinuity analysis. *Economics of Education Review*, 62, 302–320. <https://doi.org/10.1016/j.econedurev.2017.08.001>
- Jackson, C. K., Rockoff, J. E., & Staiger, D. O. (2014). Teacher Effects and Teacher-Related Policies. *Annual Review of Economics*, 6(1), 801–825. <https://doi.org/10.1146/annurev-economics-080213-040845>
- Jerrim, J., & Sims, S. (2019). *The Teaching and Learning International Survey (TALIS) 2018*.
- Ji, M. F., & Wood, W. (2007). Purchase and consumption habits: Not necessarily what you intend. *Journal of Consumer Psychology*, 17(4), 261–276. [https://doi.org/10.1016/S1057-7408\(07\)70037-2](https://doi.org/10.1016/S1057-7408(07)70037-2)
- Johnson, S., Cooper, C., Cartwright, S., Donald, I., Taylor, P., & Millet, C. (2005). The experience of work-related stress across occupations. *Journal of Managerial Psychology*, 20(2), 178–187. <https://doi.org/10.1108/02683940510579803>
- Jovanovic, B., & Nyarko, Y. (1995). A Bayesian learning model fitted to a variety of empirical learning curves. *Brookings Papers on Economic Activity. Microeconomics*, 247–305.
- Kagan, D. M. (1992). Professional Growth Among Preservice and Beginning Teachers. *Review of Educational Research*, 62(2), 129–169. <https://doi.org/10.3102/00346543062002129>
- Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment. *Research Paper. MET Project. Bill & Melinda Gates Foundation*, 1–48. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.638.2716>
- Knight, J. (2007). *Instructional coaching: A partnership approach to improving instruction*. Corwin Press.
- Kraft, M. A. (2019). Teacher effects on complex cognitive skills and social-emotional competencies. *Journal of Human Resources*, 54(1), 1–36. <https://doi.org/10.3368/JHR.54.1.0916.8265R3>
- Kraft, M. A., & Blazar, D. (2017). *Individualized Coaching to Improve Teacher Practice Across Grades and Subjects: New Experimental Evidence*. *Educational Policy* (Vol. 31). <https://doi.org/10.1177/0895904816631099>
- Kraft, M. A., Blazar, D., & Hogan, D. (2016). The Effect of Teacher Coaching on Instruction and Achievement: A Meta-Analysis of the Causal Evidence, 76. <https://doi.org/10.1086/471641>
- Kraft, M. A., & Papay, J. P. (2014). Can Professional Environments in Schools Promote Teacher Development? Explaining Heterogeneity in Returns to Teaching Experience. *Educational Evaluation and Policy Analysis*, 36(4), 476–500. <https://doi.org/10.3102/0162373713519496>
- Kraft, M. A., Papay, J. P., Chi, O. L., & Kraft, M. A. (2019). Teacher Skill Development : Evidence from Performance Ratings by Principals Teacher Skill Development : Evidence from Performance Ratings by Principals, (19). Retrieved from https://scholar.harvard.edu/files/mkraft/files/kraft_papay_chi_2018_teacher_skill_development.pdf
- Lally, P., Van Jaarsveld, C., Potts, H., & Wardle, J. (2010). How are habits formed: Modelling habit

- formation in the real world. *European Journal of Social Psychology*, 40, 998–1009.
<https://doi.org/10.1002/ejsp>
- Lashley, K. S. (1951). *The problem of serial order in behavior* (21st ed.). Bobbs-Merrill.
- Lee, S. W. (2018). Pulling Back the Curtain: Revealing the Cumulative Importance of High-Performing, Highly Qualified Teachers on Students' Educational Outcome. *Educational Evaluation and Policy Analysis*, 40(3), 359–381. <https://doi.org/10.3102/0162373718769379>
- Lehericy, S., Benali, H., Van de Moortele, P.-F., Pelegrini-Issac, M., Waechter, T., Ugurbil, K., & Doyon, J. (2005). Distinct basal ganglia territories are engaged in early and advanced motor sequence learning. *Proceedings of the National Academy of Sciences*, 102(35), 12566–12571. <https://doi.org/10.1073/pnas.0502762102>
- Leinhardt, G., & Greeno, J. G. (1986). The Cognitive Skill of Teaching. *Journal of Educational Psychology*, 78(2), 75–95. <https://doi.org/10.1037/0022-0663.78.2.75>
- Leithwood, K., Steinbach, R., & Jantzi, D. (2002). School Leadership and Teachers' Motivation to Implement Accountability Policies. *Educational Administration Quarterly*, 38(1), 94–119. <https://doi.org/10.1177/0013161x02381005>
- Lungu, O., Monchi, O., Albouy, G., Jubault, T., Ballarin, E., Burnod, Y., & Doyon, J. (2014). Striatal and hippocampal involvement in motor sequence chunking depends on the learning strategy. *PLoS ONE*, 9(8), 25–27. <https://doi.org/10.1371/journal.pone.0103885>
- Mattei, P. (2012). Market accountability in schools: Policy reforms in England, Germany, France and Italy. *Oxford Review of Education*, 38(3), 247–266. <https://doi.org/10.1080/03054985.2012.689694>
- Morrison, S. E., Bamkole, M. A., & Nicola, S. M. (2015). Sign tracking, but not goal tracking, is resistant to outcome devaluation. *Frontiers in Neuroscience*, 9(DEC), 468. <https://doi.org/10.3389/fnins.2015.00468>
- National Union of Teachers. (2008). *Teachers Working Time and Duties: An NUT Guide*. Retrieved from <https://www.teachers.org.uk/files/Workload-A5-7037.pdf>
- Neal, D. T., Wood, W., Labrecque, J. S., & Lally, P. (2012). How do habits guide behavior? Perceived and actual triggers of habits in daily life. *Journal of Experimental Social Psychology*, 48(2), 492–498. <https://doi.org/10.1016/j.jesp.2011.10.011>
- Neal, D. T., Wood, W., Wu, M., & Kurlander, D. (2011). The pull of the past: When do habits persist despite conflict with motives? *Personality and Social Psychology Bulletin*, 37(11), 1428–1437. <https://doi.org/10.1177/0146167211419863>
- OECD. (2013). *TALIS 2013*.
- Olson, M. R., & Osborne, J. W. (1991). Learning to teach: The first year. *Teaching and Teacher Education*, 7(4), 331–343. [https://doi.org/10.1016/0742-051X\(91\)90003-8](https://doi.org/10.1016/0742-051X(91)90003-8)
- Ost, B. (2014). How do teachers improve? The relative importance of specific and general human capital. *American Economic Journal: Applied Economics*, 6(2), 127–151. <https://doi.org/10.1257/app.6.2.127>
- Ostlund, S. B., Winterbauer, N. E., & Balleine, B. W. (2009). Evidence of Action Sequence Chunking in Goal-Directed Dorsomedial Prefrontal Cortex. *Journal of Neuroscience*, 29(25), 8280–8287. <https://doi.org/10.1523/JNEUROSCI.1176-09.2009>
- Ouellette, J. A., & Wood, W. (1998). Habit and Intention in Everyday Life: The Multiple Processes by Which Past Behavior Predicts Future Behavior. *Psychological Bulletin*, 124(1), 54–74. Retrieved from <http://www.duke.edu/~wwood/Ouellette.Wood.1998.pdf>
- Papay, J. P., & Kraft, M. A. (2015). Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics*, 130, 105–119. <https://doi.org/10.1016/j.jpubeco.2015.02.008>
- Perryman, J., Ball, S., Maguire, M., & Braun, A. (2011). Life in the Pressure Cooker - School League Tables and English and Mathematics Teachers' Responses to Accountability in a Results-Driven Era. *British Journal of Educational Studies*, 59(2), 179–195. <https://doi.org/10.1080/00071005.2011.578568>
- Perryman, J., & Calvert, G. (2019). What motivates people to teach, and why do they leave? Accountability, performativity and teacher retention. *British Journal of Educational Studies*, 1–21. <https://doi.org/10.1080/00071005.2019.1589417>
- Resnick, L. B. (1972). Teacher Behavior in an Informal British Infant School. *The School Review*,

- 81(1), 63–83. <https://doi.org/10.1086/443062>
- Robbins, T. W., & Costa, R. M. (2017). Habits. *Current Biology*, 27(22), R1200–R1206. <https://doi.org/10.1016/j.cub.2017.09.060>
- Roosendaal, B., McEwen, B. S., & Chattarji, S. (2009, June 13). Stress, memory and the amygdala. *Nature Reviews Neuroscience*. Nature Publishing Group. <https://doi.org/10.1038/nrn2651>
- Ruh, N., Cooper, R. P., & Mareschal, D. (2010). Action selection in complex routinized sequential behaviors. *Journal of Experimental Psychology: Human Perception and Performance*, 36(4), 955–975. <https://doi.org/10.1037/a0017608>
- Saling, L. L., & Phillips, J. G. (2007). Automatic behaviour: Efficient not mindless. *Brain Research Bulletin*, 73(1–3), 1–20. <https://doi.org/10.1016/J.BRAINRESBULL.2007.02.009>
- Schon, D. (1983). *The Reflective Practitioner: How Professionals Think in Practice*. London, Temple Smith.
- Schuler, R. S. (1980). Definition and conceptualization of stress in organizations. *Organizational Behavior and Human Performance*, 25(2), 184–215. [https://doi.org/10.1016/0030-5073\(80\)90063-X](https://doi.org/10.1016/0030-5073(80)90063-X)
- Schwabe, L., Oitzl, M. S., Philippson, C., Richter, S., Bohringer, A., Wippich, W., & Schachinger, H. (2007). Stress modulates the use of spatial versus stimulus-response learning strategies in humans. *Learning and Memory*, 14(1), 109–116. <https://doi.org/10.1101/lm.435807>
- Schwabe, L., & Wolf, O. T. (2009). Stress prompts habit behavior in humans. *Journal of Neuroscience*, 29(22), 7191–7198. <https://doi.org/10.1523/JNEUROSCI.0979-09.2009>
- Schwabe, L., & Wolf, O. T. (2010). Socially evaluated cold pressor stress after instrumental learning favors habits over goal-directed action. *Psychoneuroendocrinology*, 35(7), 977–986. <https://doi.org/10.1016/j.psyneuen.2009.12.010>
- Schwabe, L., Wolf, O. T., & Oitzl, M. S. (2010, March 1). Memory formation under stress: Quantity and quality. *Neuroscience and Biobehavioral Reviews*. Pergamon. <https://doi.org/10.1016/j.neubiorev.2009.11.015>
- Seger, C. A., & Spiering, B. J. (2011). A Critical Review of Habit Learning and the Basal Ganglia. *Frontiers in Systems Neuroscience*, 5(August), 1–9. <https://doi.org/10.3389/fnsys.2011.00066>
- Smith, A., Brice, C., Collins, A., Matthews, V., & McNamara, R. (2000). *The scale of occupational stress: A further analysis of the impact of demographic factors and type of job.* (Contract Research Report 311/2000).
- Smith, K. S., & Graybiel, A. M. (2013). A dual operator view of habitual behavior reflecting cortical and striatal dynamics. *Neuron*, 79(2), 361–374. <https://doi.org/10.1016/j.neuron.2013.05.038>
- Smith, K. S., & Graybiel, A. M. (2016). Habit formation coincides with shifts in reinforcement representations in the sensorimotor striatum. *Journal of Neurophysiology*, 115(3), 1487–1498. <https://doi.org/10.1152/jn.00925.2015>
- Snihotta, F. F., & Pesseau, J. (2012). The habitual use of the self-report habit index. *Annals of Behavioral Medicine*, 43(1), 139–140. <https://doi.org/10.1007/s12160-011-9305-x>
- Staiger, D. O., & Kane, T. J. (2015). Making Decisions with Imprecise Performance Measures. In *Designing Teacher Evaluation Systems* (pp. 144–169). San Francisco: John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119210856.ch5>
- Steele, C. J., & Penhune, V. B. (2010). Specific Increases within Global Decreases: A Functional Magnetic Resonance Imaging Investigation of Five Days of Motor Sequence Learning. *Journal of Neuroscience*, 30(24), 8332–8341. <https://doi.org/10.1523/JNEUROSCI.5569-09.2010>
- Taatgen, N. A., Huss, D., Dickison, D., & Anderson, J. R. (2008). The Acquisition of Robust and Flexible Cognitive Skills. *Journal of Experimental Psychology: General*, 137(3), 548–565. <https://doi.org/10.1037/0096-3445.137.3.548>
- Tang, C., Pawlak, A. P., Prokopenko, V., & West, M. O. (2007). Changes in activity of the striatum during formation of a motor habit. *European Journal of Neuroscience*, 25(4), 1212–1227. <https://doi.org/10.1111/j.1460-9568.2007.05353.x>
- Thorn, C. A., Atallah, H., Howe, M., & Graybiel, A. M. (2010). Differential Dynamics of Activity Changes in Dorsolateral and Dorsomedial Striatal Loops during Learning. *Neuron*, 66(5), 781–795. <https://doi.org/10.1016/j.neuron.2010.04.036>
- Toth, J. P. (1996). Conceptual Automaticity in Recognition Memory: Levels-of-processing Effects on Familiarity. *Canadian Journal of Experimental Psychology*, 50(1), 123–138.

- <https://doi.org/10.1037/1196-1961.50.1.123>
- Tricomi, E. M., Balleine, B. W., & O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *Eur J Neurosci*, *29*(11), 2225–2232. <https://doi.org/10.1111/j.1460-9568.2009.06796.x>
- Vasquez-Levy, D. (1998). Features of Practical Argument Engagement. *Teaching and Teacher Education*, *14*(5), 535–550.
- Veenman, S. (1984). Perceived Problems of Beginning Teachers. *Review of Educational Research*, *54*(2), 143–178. <https://doi.org/10.3102/00346543054002143>
- Verplanken, B., Aarts, H., Van Knippenberg, A., & Moonen, A. (1998). Habit versus planned behaviour: A field experiment. *British Journal of Social Psychology*, *37*(1), 111–128. <https://doi.org/10.1111/j.2044-8309.1998.tb01160.x>
- Verplanken, B., & Orbell, S. (2003). Reflections on Past Behavior: A Self-Report Index of Habit Strength. *Journal of Applied Social Psychology*, *33*(6), 1313–1330. <https://doi.org/10.1111/j.1559-1816.2003.tb01951.x>
- Webb, T. L., & Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychological Bulletin*, *132*(2), 249–268. <https://doi.org/10.1037/0033-2909.132.2.249>
- Wheatley, T., & Wegner, D. M. (2001). Psychology of automaticity of action. In *International encyclopedia of the social and behavioral sciences*. New York: Elsevier Science.
- Wheldall, K., Houghton, S., & Merrett, F. (1989). Natural rates of teacher approval and disapproval in British secondary school classrooms. *British Journal of Educational Psychology*, *59*(1), 38–48. <https://doi.org/10.1111/j.2044-8279.1989.tb03074.x>
- Wideen, M., Mayer-Smith, J., & Moon, B. (1998). A Critical Analysis of the Research on Learning to Teach: Making the Case for an Ecological Perspective on Inquiry. *Review of Educational Research*, *68*(2), 130–178. <https://doi.org/10.3102/00346543068002130>
- William, D. (2007). Changing Classroom Practice. *Educational Leadership*, *65*(December 2007), 36–42. <https://doi.org/10.1007/978-90-481-2660-6>
- William, L. B., & Harter, N. (1899). Studies on the telegraphic language: The acquisition of a hierarchy of habits. *Psychological Review*, *6*(4), 345–375. <https://doi.org/10.1037/h0073117>
- Williams, Z. M., & Eskandar, E. N. (2006). Selective enhancement of associative learning by microstimulation of the anterior caudate. *Nature Neuroscience*, *9*(4), 562–568. <https://doi.org/10.1038/nn1662>
- Wise, C. (2001). The Monitoring Role of the Academic Middle Manager in Secondary Schools. *Educational Management & Administration*, *29*(3), 333–341. <https://doi.org/10.1177/0263211x010293007>
- Wolfram, M., Bellingrath, S., Feuerhahn, N., & Kudielka, B. M. (2013). Cortisol responses to naturalistic and laboratory stress in student teachers: Comparison with a non-stress control day. *Stress and Health*, *29*(2), 143–149. <https://doi.org/10.1002/smi.2439>
- Wood, W., Labreque, J., Lin, P.-Y., & Runger, D. (2014). Habits in Dual-Process Models. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 371–385). Retrieved from https://books.google.co.uk/books?hl=en&lr=&id=prtaAwAAQBAJ&oi=fnd&pg=PA371&dq=Habits+in+dual+process+models&ots=ZUw_sdKzml&sig=tSgD2xmmhp-DVO13D_ce2scchXk#v=onepage&q=Habits+in+dual+process+models&f=false
- Wood, W., & Runger, D. (2016). Psychology of Habit. *Annual Review of Psychology*, *67*(1), 289–314. <https://doi.org/10.1146/annurev-psych-122414-033417>
- Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience*, *7*(6), 464–476. <https://doi.org/10.1038/nrn1919>
- Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2006). Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning. *Behavioural Brain Research*, *166*(2), 189–196. <https://doi.org/10.1016/j.bbr.2005.07.012>
- Yonelinas, A. P., & Jacoby, L. L. (1994). Dissociations of processes in recognition memory: effects of interference and of response speed. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Experimentale*, *48*(4), 516–535. <https://doi.org/10.1037/1196-1961.48.4.516>

- Yonelinas, A. P., & Jacoby, L. L. (1995). The Relation between Remembering and Knowing as Bases for Recognition: Effects of Size Congruency. *Journal of Memory and Language*, 34(5), 622–643. <https://doi.org/10.1006/jmla.1995.1028>
- Yonelinas, A. P., & Jacoby, L. L. (2012, July 18). The process-dissociation approach two decades later: Convergence, boundary conditions, and new directions. *Memory and Cognition*. Springer-Verlag. <https://doi.org/10.3758/s13421-012-0205-5>
- Zwosta, K., Ruge, H., Goschke, T., & Wolfensteller, U. (2018). Habit strength is predicted by activity dynamics in goal-directed brain systems during training. *NeuroImage*, 165(September 2017), 125–137. <https://doi.org/10.1016/j.neuroimage.2017.09.062>
-