Distribution Agreement

Signature:

________________________________ _______________
Megan E. Beney Kilgore Date

Kid Knapping: Developmental and Individual Differences in How Children and Adults Learn to Make Stone Tools

By

Megan E. Beney Kilgore
Doctor of Philosophy

Anthropology

Dietrich Stout
Advisor

Melvin Konner
Committee Member

Lynne Nygaard
Committee Member

Accepted:

Kimberly Jacob Arriola, Ph.D, MPH
Dean of the James T. Laney School of Graduate Studies

Date

Kid Knapping: Developmental and Individual Differences in How Children and Adults Learn to Make Stone Tools

By

Megan E. Beney Kilgore
B.A. Anthropology and Music, Knox College, 2013

Advisor: Dietrich Stout, Ph.D.

An abstract of
a dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Anthropology
2023

ABSTRACT

Kid Knapping: Developmental and Individual Differences in How Children and Adults Learn to Make Stone Tools

By: Megan E. Beney Kilgore

This dissertation investigates the cognitive, motor, and personality differences in how children and adults learn how to make least effort, Oldowan-style technology in an experimental stone toolmaking task. Child participants (N = 19, ages 8 – 13 years, mean = 9.3 years) were recruited in collaboration with the Tellus Science Museum (Cartersville, GA) as a part of the organization's summer educational programming. In addition to their optional participation in the dissertation study, children engaged in science education outreach activities that explored the discipline of archaeology and took advantage of the museum's curated galleries and classroom resources. Adult participants (N = 30, ages 18 – 38 years, mean = 20.3 years) were recruited to the Paleolithic Technology Laboratory at Emory University (Atlanta, GA). All participants, regardless of age, were asked to complete five consecutive days of stone toolmaking training in 30-minute sessions, for a total of 2.5 hours of practice. Participant group sizes varied from one-on-one to groups of up to five, depending on participant availability. All participants also took a series of psychometric and motor testing related to grip strength ability, executive functioning, motor control and coordination, and personality. Measurements taken on the cores and lithic debitage yielded three PCA factors, interpreted as the knapper skill outcome variables Quantity, Quality, and Flaking Inefficiency. Results indicate that adults routinely outperformed children both in psychometric and motor performance and in knapper skill outcomes. However, adult male participants distinguished themselves from all other participant subgroups (i.e. adult females and male and female children) in that they produced lithics of the best Quality, a result that may be tied to their superior grip strength. These findings contribute to the broader literature on experimental archaeology and provide a unique developmental perspective on stone toolmaking skill acquisition.

Kid Knapping: Developmental and Individual Differences in How Children and Adults Learn to Make Stone Tools

By

Megan E. Beney Kilgore
B.A. Anthropology and Music, Knox College, 2013

Advisor: Dietrich Stout, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Anthropology
2023

ACKNOWLEDGEMENTS

They say it takes a village to raise a child. With the recent birth of my son and my entry into motherhood, I can attest to the phenomenal social support required for individuals who undergo great labors. This dissertation is one such labor, and there are many people who helped make its completion possible.

First and foremost, I must thank my advisor, Dietrich Stout. He has guided me through the many twists and turns of my graduate experience, and his patience, unwavering support, dry wit, and remarkable tolerance of my attempts to transform the lab into a cozy space have been the cornerstone of my time at Emory. I am eternally grateful.

I also offer my many thanks to my committee members, Melvin Konner and Lynne Nygaard. Mel, your constant kindness and gentle words of encouragement, especially when I was doubting myself, helped carry me through my graduate experience more than you know. Lynne, becoming an unofficial Laabie was a highlight of my time at Emory, and I hope to carry out all my future work with a sense of perspicacity and aplomb, in your honor.

My sincerest gratitude goes to the anthropology administrative staff, past and present, for their timely and helpful replies to my many questions, for their generous feedback to my baking efforts, and for their friendship. I would like to especially acknowledge Lora McDonald, Eva Stotz, Heather Carpenter, Morgan Wilson, and Brian Banks. All of you have been indispensable to me, and I am forever in your debt.

I would like to also acknowledge my funding sources, the Emory Professional Development Support (PDS) funds of Laney Graduate School and the National Science Foundation (NSF) for their award of a Doctoral Dissertation Research Improvement Grant. Without their financial support, this project would not have been possible.

To my friends and colleagues at the Tellus Science Museum, I do not exaggerate when I say that without your support, this dissertation would never have come to fruition, especially in a time of global pandemic. Special thanks go to David Dundee, Hannah Stewart, and Wendy Hayes for their administrative guidance and to Kady Yeomans for keeping me out of, or getting me into, the best kind of trouble. I also want to thank the many Tellus education staff whose enthusiasm and hard work turned our vision of Archaeology Adventure Week into a reality.

I am grateful to my cohorts, both in my home department of anthropology and in the CMBC certificate program. Special thanks go to Minwoo Lee, who existed in both these spaces, for our many thoughtful conversations and shared love of cats. I also want to thank my friends outside of my own cohort, especially Grace Veatch, Kaitlin Banfill, SJ Dillion, Sophie Joseph, and Bre Meyers. Our game nights, coffee/tea breaks, and hallway conversations meant the world to me.

My most heartfelt thanks go to my lab mates, Cheng Liu and Aditi Majoe. Cheng, you have been my most stalwart companion, and despite your aversion to sentimentality, you should know how much I've valued your constant encouragement, good company, and willingness to put up with my bouts of silliness. Aditi, our shenanigans, inside and outside of the lab, comprise some of my fondest memories at Emory.

To my many research assistants – Atlas Moss, Valeria Funes, William Yang, Noah Larco, Yulia Gu, Krithika Shrinivas, and Madina Eshova – I am grateful for your hard work, willingness to learn, and for the opportunity to transform our little group into a Lab Family. May your warm beverages always be at perfect sipping temperature and your Starbursts never be orange.

My utmost love and thanks go to my friends and family: To my mother, who has always believed in me, more than I believed in myself. To my dad and step-mom, who are among my greatest champions. To my brother, whose computer wizardry has gotten me out of many a bind. To Madison, my sister by heart, who is both my partner in crime and biggest advocate.

To my husband, Matthew, you have kept me going, mind, body, and spirit. If this dissertation belongs to anyone, it belongs to you.

Lastly, to my son. Cedrick Benjamin, my storm-born boy, your early arrival into this world was a glorious disruption to the final stage of my dissertation. You are my greatest new adventure, and perhaps we'll follow in the illustrious footsteps of Piaget, whose observations of his own children inspired his theories of child development, and teach you how to make stone tools.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

CHAPTER 1: INTRODUCTION AND THEORY

INTRODUCTION

That children existed in and contributed to prehistory is undisputed; however, despite this obvious claim, the task of identifying children's presence in the archaeological record remains a challenging one. Although the last few decades have seen an influx of interest in this topic (Lillehammer 1989; Baxter 2008; Shea 2006; Finlay 1997) much of the existing literature focuses on high level theory and/or attempts to identify children in archaeological assemblages, with little effort directed toward actualistic work that could put both endeavors on a firmer foundation. While this literature has been crucial in legitimizing the archaeology of childhood to the general research community, the paucity of empirical studies has made it difficult to address broader implications for hominin biological and cultural evolution. To achieve this, it is critical to move beyond exclusive usage of intuitive criteria for identifying children in the archaeological record into research that both establishes empirical standards for identifying ancient children's work and addresses the behavioral, cognitive, and social aspects of child material culture.

Despite its clear importance to human evolution, scholarly work on stone toolmaking in children is scarce, with only one empirical study published on the subject, to date (Sternke and Sørensen 2005). The inclusion of a developmental perspective on social learning and stone toolmaking is important for several reasons. First, from the basis of hominin life history, it is almost certain that children would have been the targets of social transmission for stone toolmaking skill acquisition. As with children in extant forager societies (Boyette 2013; Hewlett and Roulette 2016; Boyette and Hewlett 2017; Crittenden 2016; Konner 2016), ancient hominin children likely would have started learning skilled behaviors related to subsistence around the time of middle childhood (i.e. ages 7 – 12 years old). A uniquely human life history stage, there is evidence that a distinct childhood phase emerged in hominin development as early as *Homo habilis* (Bogin 1997), or ~2.4 – 1.4 million years ago (mya). This time period coincides with the known dates for the Oldowan Stone Industry (~2.6 – 1.2 mya), so it is not only possible but probable that ancient hominins would have learned to make stone tools as children.

Second, this question is fundamental to our understanding of stone toolmaking skill acquisition given that individuals at varying stages of development have different cognitive and motor abilities. Namely, because the motor and cognitive abilities of children differ from those of adults, we might expect there to be differences in the amount and types of teaching necessary to transmit the requisite skills for stone toolmaking. Furthermore, the reasons behind differences in children's and adults' approaches to skilled behaviors, such as foraging, are not entirely understood. Though it may be easy to dismiss these differences as evidence only of children's relative lack of development and inexperience, they may be significant in the sense that, in some cases, being unskilled provides a paradoxical cognitive advantage to learners. In other words, pre-existing knowledge and ability can sometimes disadvantage someone trying to learn a new skill. The differences in how children and adults learn are well-documented (Kuhn and Pease 2006; Du et al. 2017), and if they are demonstrated in stone toolmaking, it will be important to revisit the conclusions draw by previous literature on this subject, particularly those that discuss the broader evolutionary context of social learning and language in stone toolmaking.

In terms of motor performance, there is some debate regarding whether children underperform compared to adults on skill-based tasks because they are simply incapable of matching adult outcomes due to a lack of gross and fine motor development (Gurven, Davison, and Kraft 2020) or because their performance is optimized for their particular body size and shape and therefore looks inherently different from adult performances (Bird and Bliege Bird

2000, 2002). The source of this difference is relevant to how both ancient hominin and modern human children may approach the task of stone toolmaking and whether it is justifiable to compare child and adult stone toolmaking on a one-to-one basis. If it is demonstrated that children and adults have significantly different learning and motor approaches to stone toolmaking, this will have implications for past conclusions on Paleolithic social learning processes and on how research on this topic is conducted in the future.

Hominin Evolution and The Origins of the Modern Mind

The uniqueness of hominin cognition has long been a focus of anthropological research. Increases in hominin brain size and the corresponding behavioral complexity demonstrated in the archaeological record (i.e. via stone toolmaking (Ambrose 2001; Shea 2017), cultivation of fire (Sandgathe 2017), and evidence of symbolic culture through art and artifacts (Dissanayake 2017; Sterelny 2017) are often pointed to as hallmarks of our genus (Antón and Snodgrass 2012). Despite these broadly agreed-upon milestones, methods for investigating the evolution of hominin cognition remain diverse and, at times, in conflict with one another.

In the earliest days of studying ancient hominin cognition, researchers relied on methods such as the use of endocasts of hominin crania – which provided basic information about a brain's general shape and, if lucky, the convolutions of its sulci and gyri – and a derived numeric "cranial capacity" as a stand-in for brain size (Falk 2012). Proponents of the former argued that endocasts are the only direct method of studying brain evolution, as they are made from imprints of actual fossilized cranial specimens and are therefore researchers' exclusive avenue into what ancient brains may have looked like and how they were structured. Referred to as "paleoneurology," the use of an endocast was particularly influential in Raymond Dart's (1925) interpretation of the Taung child (a juvenile *Australopithecus africanus*) as an early, ape-like ancestor of the genus *Homo*. This method, however, has been met with considerable criticism, as the data provided by endocasts is relatively limited and at risk of overinterpretation, although with the introduction of new technology, notably computerized tomography (CT) scans and geometric morphometric techniques, some researchers have called for a re-examination of the endocast's usefulness in the study of ancient hominin cognition (Neubauer 2014).

The latter method, calculating a fossil specimen's cranial capacity as a proxy for brain size, has remained a somewhat more popular approach to the study of ancient hominin cognition. Although this method also suffers from its share of inaccuracies (Holloway 1966), it has given rise to a body of literature on the comparative cognition of humans and extant primate species and uses fossilized ancient hominin crania, instead, as a reference point for arguments on cognitive evolution, rather than deriving conclusions about ancient hominin behavior from fossils, alone. This combination of techniques has resulted in a more robust research program, from which have emerged theories on the roles of encephalization quotient (Jerison 1979; Vančata 1996); brain size (Gonzalez-Forero and Gardner 2018; Navarrete, van Schaik, and Isler 2011; Shultz, Nelson, and Dunbar 2012), growth (Leigh 2004), and reorganization (Hecht et al. 2013; Holloway, Broadfield, and Yuan 2003; Smaers and Soligo 2013); and social (Dunbar 1998), behavioral (Burkart, Hrdy, and Van Schaik 2009; Byrne 2000; Sterelny 2011), and ecological (Burkart, Schubiger, and van Schaik 2017) complexity on the evolution of hominin cognition. These efforts have produced a mosaic-like picture of the processes that contributed to and shaped hominin cognition, where many variables worked in tandem to create the modern human mind.

Behavior, Grounded Cognition, and Language

Behavior as an Evolutionary Agent

Of these variables, the study of behavior's role in evolutionary processes has experienced a resurgence over recent decades. In 1896, following the frenzy of scientific thought inspired by Darwin's 1859 publication of *On the Origin of Species*, psychologist James Mark Baldwin postulated that, under certain conditions, learned behaviors can impact the direction and thrust of a species' evolutionary trajectory through natural selection (Simpson 1953). More specifically, as a learned behavior (e.g. tool use) proved itself to be adaptive, individuals who adopted this behavior would experience increased survivability, and the traits associated with successful application of the behavior (e.g. executive function) would undergo the process of natural selection, should the behavioral practice persist over evolutionary time, and remain present, if not more prominent, in those individuals' descendants. This concept, aptly named the "Baldwin effect," fell out of favor in the early and mid-twentieth century because of its perceived proximity to neo-Lamarckian positions on evolutionary theory and concerns surrounding the difficulty of empirically proving its existence (Depew 2003).

Since then, the Baldwin effect, though not always mentioned explicitly, has found new traction in research on hominin life history and the Extended Evolutionary Synthesis (EES). Championed by Kevin Laland and John Odling-Smee, among others, the EES proposes a framework that more formally integrates processes usually perceived as proximate to genetic inheritance (i.e. developmental bias, developmental plasticity, inclusive inheritance, and niche construction) into a view of evolution that emphasizes the constructive development of gene-environment interactions and reciprocal causation between species, behavior, and environment (Laland et al. 2015). At its core, the EES states that the synergistic effects of organism and environment may function to bias selection, mediated by the plasticity of ontogenetic processes, and in doing so, species' behaviors, including those that are learned, can impact phenotypic evolution and, ultimately, genetic inheritance.

Grounded Cognition, Affordances, and the Interconnectedness of Brain, Body, Behavior, and Environment

This interconnectedness of brain, body, behavior, and environment is also emphasized in the study of embodied cognition and affordances, related concepts that, while not directly tied to evolutionary theory, can provide context for the mechanisms by which the inclusive inheritance of the EES may occur. Namely, research on grounded, or "embodied," cognition reframes the traditionally held belief that the brain exists and functions as an information processor distinct from its somatic housing. From this perspective, hominin cognition is deeply rooted in sensorimotor processing – as opposed to resting solely in a separate semantic memory system – and this processing, in turn, evolved in response to environmental input (Barsalou 2008; Wilson 2002). Furthermore, within a grounded cognition framework, minds can essentially outsource some of their processing load to their bodies and environments by relying on systems such as visual imagery, simulation, and affordances provided by physical objects and spaces (Clark 1999).

The latter of these, an affordance describes an agent's potential interactions with its environment, provided by the physical features of space and individual objects (e.g. the flatness of solid ground) and by the agent's ability to perceive these features (e.g. both in terms of cognitive resources and the availability of visible light). A term coined by J.J. Gibson (1979), he notes that affordances must be considered with respect to the individual in question, as differences in anatomy (e.g. hand size in children versus adults) will affect how the individual is able to engage with an object. In his initial publications on the subject, Gibson leaves his criteria for what constitutes an affordance somewhat broad, ascribing the term to physical, cognitive, and social phenomena. Although this generality encompasses Gibson's assertion on the ubiquity of affordances, it has resulted in a muddying of how an affordance should be defined and operationalized, especially in the literature surrounding tool use (Osiurak, Rossetti, and Badets 2017).

Authors Osiurak, Rossetti, and Badets (2017) have since revisited the properties of affordances and have highlighted the relationships between hand and tool (hand-tool) and tool and object (tool-object) as especially relevant to this literature. They furthermore contextualize these relationships in what they describe as the three action-system (3AS) model, wherein they emphasize the role of specific cerebral systems involved in an agent's perception of affordances (i.e. dorso-dorsal), the comprehension of mechanical action (i.e. ventral-dorsal), and specifying specific contextual relationships between tools and objects (i.e. ventral). Although the coupling of specific affordance qualities to cognitive systems is not new (Norman 2002; Randerath et al. 2011), it does provide an empirically grounded basis for the relationship between brain, body, behavior, and environment, one that emerges in infancy (Byrge, Sporns, and Smith 2014; Smith and Gasser 2005) and is seen both in individual's interactions with physical objects and symbolic abstractions, such as mathematics (Gibson 1979) or language (Kaschak and Glenberg 2000).

Embodied Language and Sound Symbolism

Observed first in the broad sociolinguistic framework of the Sapir-Whorf hypothesis (Boroditsky 2001; Kay and Kempton 1984; Whorf 1988) and later in the study of metaphor usage (Lakoff and Johnson 1980), the role of language in grounded cognition is multifold. Not only does it provide a shared frame of reference for interlocutors (Hasson et al. 2012), but it also acts as an anchor for ‘thinking about thinking’ or a ‘cognitive super niche’ (Clark 2006). For example, Dehaene (2011) suggests that in order to move beyond the basic, easily discernable numerical categories of oneness, twoness, and threeness, language is not only useful but necessary. More specifically, humans can, without effort, count objects in groupings of up to three; however, once groups reach sizes larger than these, they fall into the category of ‘more-than-that-ness.’ In order to talk, or perhaps even think, cohesively about larger numerical values requires a lexicon that accommodates these efforts, hence, in the case of counting and numbers, language facilitates an otherwise inaccessible cognitive process.

Historical definitions of language cite its arbitrariness as one of its hallmark characteristics (Hockett 1960; de Sassure 2013), however, there is growing evidence for the non-arbitrariness of word form-to-meaning mapping. This work on sound symbolism emphasizes the roles of iconicity (e.g. onomatopoeia, ideophones) and semanticity (e.g. prosodic or phonological cues, such as syllable length) in language structure (Dingemanse et al. 2015), both of which take advantage of the brain’s perceptuomotor system. A famous example of this process is found in the bouba-kiki effect (McCormick et al. 2015) or the maluma-takete effect (Sidhu et al. 2021), which both demonstrate participants’ tendencies to associate rounded vowel sounds (e.g. /oʊ/ and /ɑ/ in “bouba” or /ɑ/ and /u/ in “maluma”) with ameboid/blob-like shapes and sharp consonant sounds (e.g. /k/ in "kiki” or /t/ and /k/ in “takete”) with pointed/angular shapes. Not only are such associations found with the shape of objects, but they are also observed with object size (Shinohara and Kawahara 2010; Tsur 2006) and distance (Rabaglia et al. 2016), in color luminance and saturation (Johansson, Anikin, and Aseyev 2019), and across diverse language families (Blasi et al. 2016). This “groundedness” of sound symbolism in language is also apparent in studies of grasp affordances, wherein participants who are asked to simultaneously produce the front-closed vowel /i/ while performing a precision grip and those who were asked to produce the back-open vowel /ɑ/ while performing a power grip responded more quickly than when asked to perform the opposite pairing (Vainio et al. 2019). The interconnectedness of language, brain, and body, well-established in the sound symbolism literature, acts as a compelling primary example of grounded cognition at work in the hominin lineage.

Summary

Taken together, the role of behavior in evolutionary processes, the embodiment of the mind, and the non-arbitrariness of sound symbolism in language work to create an image of hominin cognitive evolution that is intimately interwoven with its context, where the human brain could not exist as it is today without the behaviors, bodies, or environments of hominins past. One specific such behavior, often cited as integral to hominin evolution, is the creation of stone tools.

Stone Toolmaking and the Evolution of Cognition and Language

Early Stone Industries

As one of the oldest sources of evidence for early hominin material culture, stone tools provide an avenue into the study of ancient cognition and behavior (Wynn and Coolidge 2016). Several Lower Paleolithic (LP) stone industries have been identified and defined, the oldest being the contested Lomekwian (~3.3 mya), followed by the Oldowan (~2.6 – 1.2 mya), Acheulean (~1.7 – 0.2 mya), and finally, the Levallois (~0.3 mya), a production technique that begins in the LP but is most often associated with the Mousterian culture of the Middle Paleolithic. Of these, the Oldowan and Acheulean have received the most attention in studies of ancient hominin cognition, owing to their pervasiveness in the archeological record and their cumulative ~2.4 million years of hominin life history.

Although, historically, Oldowan technology was described by a typology of core forms (e.g. "choppers," "scrapers," and "discoids") (Leakey 1971), more recently, the Oldowan has been recognized as a core and flake industry (Toth 1985). It is produced when a stone knapper uses one stone (the hammerstone) to strike another stone (the core) to create a sharp, knife-like fragment (the flake). Where the knapper chooses to strike the core is of great importance, as surfaces (platforms) with acute angles are more likely to result in flakes with sharper, more viable cutting edges. The ability to make this choice relies on the toolmaker's capacity to recognize the angle affordances that the core provides, in addition to visuomotor skill and a basic understanding of fracture mechanics (Stout and Chaminade 2007). A range of hominin species are contemporaneous with Oldowan archaeological assemblages, dating back to *Paranthropus aetheopicus* and *Australopithecus garhi*, but the most compelling associations between Oldowan tools and toolmakers lie with *Paranthropus boisei*, *Homo habilis*, and *Homo erectus* (Toth and Schick 2018). The cranial capacity of the former two hominin species is modest, at 500-550cm^3 and 510-690cm^3, respectively. *Homo erectus*, who was also alive to at the correct time to manufacture Acheulean technology, boasted a cranial capacity of 800-1070cm^3, the upper range of which overlaps with the lower range of modern humans.

Acheulean technology is marked by several key features, called "imperatives" by Gowlett (2006): the glob-butt (the rounded proximal end of the handaxe, serving as its center of gravity), forward extension (counterbalancing the glob-butt), lateral extension (offering resistance to torque and providing support for long working edges), support for the working edges (in relation to the glob-butt), and thickness adjustment (allowing for the general mass to be worked without much loss to the rest of the piece). Unlike Oldowan flakes, which are typically produced unifacially, Acheulean handaxes are created through the bifacial removal of flaked pieces along a central, lateral axis to create a cutting edge. Knappers exploit existing flake scars to assist in the shaping and thinning of the worked tool, following what appears to be a prescribed planning sequence, which again differentiates Acheulean stone toolmakers from the Oldowan, who showed no clear evidence of a mental template in their own reduction sequences (de la Torre 2016).

Methods for Studying Ancient Cognition in Toolmaking

Because the ancestral hominin mind is unknowable through modern psychological methods, given that none of its agents are still alive, it is important to take a multidisciplinary approach to the study of ancient cognition, both through a series of linked inferences (Coolidge and Wynn 2016) and experimental work conducted with modern human participants. The latter of these I will discuss in greater detail below; the former is achieved through a sequence of bridging arguments between what *is* known about the past (i.e. through lithic artifacts) and subsequent conclusions drawn from this knowledge (e.g. categories of technological industries, methods of reducing stone cores into useable tools). These methods allow researchers to create what is called a *minimal-capacity inference*, wherein they discern what sort of preconditions (e.g. material, cognitive, motor) are necessary to create an object in question – in this case, stone tools – and then infer these capacities to the object's maker (Currie and Killin 2019). Although researcher reliance on pure logic to address evolutionary questions is mitigated, in part, by the advent of new technologies (e.g. radio carbon dating), certain aspects of the evolutionary past, such as stone toolmaker (knapper) intention and identity and ancient hominin cognitive ability, remain obscured from testable, empirical methods and still require more orthogonal means of inquiry, such as through archaeological analogy (Ascher 1961).

In the case of hominin cognitive evolution and its relationship to stone toolmaking, work in cognitive archaeology has demonstrated that stone knappers make their tools through a series of intentional, goal-oriented actions (Wynn 2002), beginning with material acquisition and transportation (Braun et al. 2008) and culminating in planned reduction sequences (de la Torre and Mora 2018; Toth and Schick 2018). In order to be successful at their task, the toolmakers would need to employ certain cognitive resources, including various aspects of executive functioning (e.g. working memory) and visuospatial reasoning, all of which can be observed in modern human participants through well-established psychological methods. The degree to which these cognitive traits played a role in stone toolmaking has been contested and likely varies across stone industries, in correspondence with the abilities of the hominin species who acted as the primary toolmakers. Namely, Oldowan flake and core technology has been cited as less cognitively demanding than the Acheulean and Levallois industries that followed it, as the latter require more advanced forms of planning to properly execute the desired end-form tool. However, while some researchers contend that Oldowan stone toolmaking requires little more than ape-like cognitive ability (Wynn et al. 2011), others promote a more medial position and assert that ancient Oldowan toolmakers had already undergone some measure of humanlike cognitive evolution (Toth and Schick 2018).

Whether language played a role in the transmission of technical skill in these various industries is also a matter of debate, with some researchers citing experimental evidence in favor of an early emergence and adoption of language in stone toolmaking knowledge transmission (Morgan et al. 2015) and others in opposition of it (Putt et al. 2017). To draw these conclusions, researchers have utilized social transmission chain and brain imaging methodologies, sometimes separately and sometimes in combination. In studies of social transmission chains, novice participants are typically presented with two or more learning conditions of stone toolmaking. At minimum, the research design will include a dichotomy of an instructed and an uninstructed condition, though these terms are variably defined. For example, sometimes "instructed" is meant to convey direct-active teaching (Kline 2016) and other times it depicts verbal instruction, only, without the use of accompanying hand gestures or other instructional modalities. Similarly, "uninstructed" conditions can mean a variety of things, including a reverse engineering condition where no instructor is present or gesture-only, nonverbal communication provided by an experimenter. Some studies provide more elaborate learning conditions, such as Morgan et al. (2015), who included reverse engineering, imitation/emulation, basic teaching, gestural teaching, and verbal teaching in their study design.

For braining imaging studies, a diversity of imaging techniques, including positron emission topography (PET) scans (Stout et al. 2000; Stout, Toth, and Schick 2006; Stout et al. 2008), functional near-infrared spectroscopy (fNIRS) (Putt et al. 2017; Putt, Wijeakumar, and Spencer 2019), magnetic resonance imaging (MRI) (Hecht et al. 2015; Stout et al. 2015), and functional magnetic resonance imaging (fMRI) (Stout et al. 2011) have been employed. These studies typically investigate the differences in skill level between novice and expert knappers, sometimes including a third, intermediate "trainee" category, and through these comparisons, they have identified regions of the brain significant to stone toolmaking in different lithic industries (i.e. the Oldowan and the Acheulean). Such regions of interest (ROIs) for Oldowan knapping include the rostral and dorsal intraparietal sulcus and ventral premotor cortex (sensorimotor control and analysis of the three-dimensional shape and orientation of the core), the medial premotor cortex (responsible for bimanual coordination), and the middle and inferior occipital gyri (visual control of action and perception of objects) (Stout and Chaminade 2007). Notably, initial brain imaging of novice Oldowan toolmaking did not reveal any significant activation in the dorsolateral prefrontal cortex, which suggests that Oldowan flake production does not require resources from executive functioning, such as working memory and inhibition control. This result is supported in a later study of expert Oldowan knapping, where it is suggested that the distinction between novice and expert Oldowan toolmakers is found, instead, in the increased motor control and the enhanced "body + tool system" of the latter (Stout et al. 2008).

The cognitive demands for Acheulean knapping reflect the same ROIs as found in the Oldowan, though to a more intense degree, in addition to areas including the right inferior/ventrolateral prefrontal cortex (responsible for coordinating flexible, goal-oriented behavior), indicating that the shift from Oldowan to Acheulean toolmaking relied on an increased ability to engage in hierarchical planning sequences of flake removal. Notably, some of these ROIs, namely the right ventral premotor cortex, which is shared with Oldowan toolmaking, overlap with those used during language production and processing. This observation, in addition to their shared emphasis on hierarchical patterning (i.e. of action, in stone knapping, and of grammar, in language), has led researchers to propose an evolutionary connection between the emergence of language and stone toolmaking in the hominin lineage.

Technological Origins of Language Hypothesis and Cumulative Culture

These brain imaging data, in combination with the social transmission studies that promote language's early influence on the transfer of stone toolmaking knowledge, corroborate what is known as the technological origins of language (TOoL) hypothesis. As the name implies, this hypothesis on the evolutionary origin of language states that language may have co-evolved with stone toolmaking as a means of achieving higher fidelity cultural transmission of skilled knowledge. TOoL is one of many hypotheses on language evolution, where others have suggested that language evolved to address any number of needs: to facilitate coordination in hunting (Washburn and Lancaster 1968), as a form of social grooming (Dunbar 1998), to soothe infants (Falk 2009) or to promote pair bonding (Deacon 1997), to name a few. However, while language would have certainly helped to facilitate any one of these scenarios, the TOoL hypothesis derives some of its strength from its shared roots in cumulative culture evolution and social learning, processes that are acknowledged as foundational to the early shaping of hominin behavior (Migliano and Vinicius 2022; Richerson and Boyd 2005).

Specifically, Laland (2017) asserts that language evolved to teach kin. The term "teaching" is variably defined and what constitutes this behavior is a moving goalpost, depending upon researcher goals and motivations. Broadly, teaching belongs to the greater behavioral family of social learning, generously defined as any behavior that can be learned through observing or imitating others. Instances of social learning are numerous and have been demonstrated in species across a wide range of taxa. Some notable cases include tandem running in ants (Leadbeater, Raine, and Chittka 2006), the waggle dance of the honeybee (Wenner 1962), sponge tool use in dolphins (Mann et al. 2012), tool design and manufacture in New Caledonian crows (Hunt and Gray 2003), and a host of behaviors in our closest living relative, the chimpanzee, such as termite fishing, nut cracking, and a characteristic grooming handclasp (Moore 2013). However, as impressive as these behaviors may be, none of them fit the definition of teaching, proper, as put forth by Caro and Hauser (1992): that, to qualify as teaching, a behavior must (1) only take place in the presence of a naïve observer, (2) come at a cost/provide no immediate benefit to the demonstrator, and (3) allow the observer to acquire knowledge or learn a skill more rapidly or effectively than it would otherwise, or that the observer would not be able to recreate the behavior in their lifetime without the aid of the demonstrator.

This definition, sometimes criticized as being too restrictive (Byrne and Rapaport 2011), is useful in that, in this last criterion, it touches on what is known as the "cultural ratchet" effect (Tennie, Call, and Tomasello 2009). Like a mechanical ratchet, whose mechanisms consist of a set of angled teeth such that, when a cog or tooth engages, motion is allowed in only one direction, Tomasello and his colleagues have described human culture as an iterative and additive process with minimal loss between generations. This is distinct from other species' social learning behaviors, even those defined as being markedly "cultural," such as the famous example of sweet potato washing in Japanese macaques (Kawai 1965), given that human cultural phenomena tend to have a high rate of survivability, even past the death of the innovating individual. Such robustness does not appear to be the case in nonhuman cultural systems, who are at risk of the extinction of innovative behaviors, and one of the causes of this difference may be attributable to the evolution of language in the hominin lineage.

The extent to which language – and by proxy, direct-active teaching – assists in the acquisition of different types of skilled knowledge is up for debate, as learning conditions for gross motor tasks, for example, may vary from fine motor tasks, which may vary from memorization tasks of semantic knowledge, and so on. This difference is demonstrated in the conflicting conclusions regarding the usefulness of language in teaching Oldowan-style stone knapping (e.g. Lombao, Guardiola, and Mosquera 2017 vs. Putt et al. 2017), whose mastery relies more on motor control proficiency than advanced executive functioning, in contrast with its clear benefit to instructing the more cognitively demanding, action-planning oriented Acheulean-style stone toolmaking (Stout et al. 2008). Furthermore, the environments and conditions in which naïve individuals learn are also variable and have undergone substantial revision within the past century. More specifically, Western-style classrooms, which emphasize learning by lecture and often feature large student-to-teacher ratios, are now more ubiquitous than ever, whereas historically, this kind of formal education was a luxury reserved for the privileged few of the upper class. Even apprenticeship-style learning, which features an extended one-on-one learning period between an expert "master" and novice "apprentice," is an invention of recent centuries and may not be representative of the learning conditions of the evolutionary past. With these caveats in mind, and acknowledging, again, that modern human subjects are also anatomically different than their ancient hominin counterparts, it is important to remember to place the conclusions from experiments such as those described above within their appropriate context. This is not to say that such efforts are futile – indeed, advances in neuroarchaeology, experimental archaeology, and their related fields give researchers the best fighting chance possible to glean information about processes, such as language origins and their relationship to stone toolmaking, integral to hominin evolution.

Ethnography of Stone Toolmaking

One other methodological avenue available to researchers of this discipline is the use of ethnography. This approach comes with its own set of fine print, namely that living human populations cannot, nor should be, viewed as a direct corollary of ancient hominin behaviors. Additionally, the behavior in question, stone knapping, has almost entirely gone out of use outside of small artisan groups that uphold the traditional practice, as other technologies and materials have replaced it. Thus, most stone knapping today is performed by hobbyists and academics (e.g. Whittaker 2004). However, one ethnography, put forth by Stout (2002) details the manufacture of adzes by both skilled and unskilled craftsmen of the Langda village in Indonesian Irian Jaya. Here, he describes the social and technological contexts for the creation of the adzes, from raw material acquisition to tool production. Of particular note are the intricate terminology used to depict the process of stone toolmaking (e.g. *ya-winwin* for a particular type of hammerstone and *temena*, for the process of roughing-out stone blanks) and the collaborative nature of training apprentices. These novices, who, according to Stout, would have traditionally begun their training around the age of 12 or 13 years old, were typically young men in their early twenties at the time of his writing. The length of an apprentice's training would span years, and the toolmaking activities and resources available to them were limited according to their level of skill. However, because adze production was a highly social process, conducted in groups of mixed expertise, novices were guided by the feedback of more experienced individuals and had access to more advanced steps of manufacture by observing and discussing the procedure with their mentors.

This form of scaffolded learning, which situates learners in what Vygotsky (1978) calls the "zone of proximal development," is an example of how skilled knowledge of stone knapping is transmitted from expert to naïve individuals. It is heavily dependent on the use of language and direct-active teaching, as well as observational learning on the part of the novice. Whether this type of learning scenario was characteristic of stone knappers of the evolutionary past is unknowable, but ethnographic work, such as that by Stout (2002), does have the potential to shed light on how modern humans may approach the task of stone toolmaking. This method, in combination with neuroarcheology, discussed above, and experimental archaeology, discussed below, can allow researchers to triangulate upon the conserved cognitive and behavioral components of stone toolmaking, so that the present may serve as a window into the past.

History of Experimental Studies of Stone Toolmaking

The study of stone toolmaking has a long history, dating back centuries, with the earliest scholarship classifying stone tools within the general category of fossils (de la Torre 2011). It was only at the end of the seventeenth century that researchers began to think of stone tools as human products. After this initial shift, archaeological perceptions of stone tools – how they were made, who made them, and the importance of concepts such as the cultural and temporal contextualizing information of archaeological assemblages and knapper knowledge, intent, and skill – all came into question, and, correspondingly, several key methodological and ideological traditions arose from the research that followed.

In the mid-twentieth century, typological approaches to studying stone tools were predominant, influenced by the work put forth by Bordes (1961), which included a detailed classification system of Mousterian artifacts in France. From this method came two of the seminal works on the early African stone industries: Leakey (1971), which described Beds I and II at Olduvai Gorge; and Isaac's (1977) report on Olorgesailie. While this approach provided a formalization of the Oldowan and Acheulean stone industries and emphasized inter-site variability and the emergence of cultural traditions in stone toolmaking, it lacked in its ability to place artifacts within the context of their specific site (e.g. through site formation processes) (de la Torre and Mora 2009). Reflecting on this shortcoming, Isaac revised his methodological approach and proclaimed that stone artifacts should be studied both in their original contexts and in the overarching framework of hominin behavior. With this declaration of purpose came the processualist worldview of archaeological research, and within processualism, experimental archaeological methods began to build traction.

In the decades that followed, much of the experimental archaeology of stone toolmaking was guided by the intuitions of expert knappers, to the detriment of robust empirical methods (Eren et al. 2016). More specifically, criteria for knapping skill and 'good' or desirable lithic products were determined from researchers' own experience with the craft, at the cost of hypothesis testing, replicability, and method validation, all of which are crucial for building and maintaining an experimental paradigm that can address with confidence questions that span archaeological space and time. Efforts to fill this methodological gap have been numerous, and our knowledge of phenomena such as the properties of stone fracture mechanics (e.g. Dibble and Whittaker 1981, Odell 1981), reduction strategies and use of design space (e.g. Newcomer 1971, Moore 2011), lithic tool function, use, and efficiency (e.g. Crabtree and Davis 1968, Tringham et al. 1974), and knapping skill acquisition (e.g. Muller, Shipton, and Clarkson 2022) has grown considerably from this scholarship. However, despite this progress, the field of experimental archaeology still suffers from its share of blind spots, most notably in the number and types of participants recruited and in experimental task training times.

Trends in Participant Recruitment and Demographics

It is no secret that the history of Western scholarship has a bias in favor of cis-gendered, adult white men from middle to upper class backgrounds, both in terms of those performing the research and those participating in it. Treated as a model for all of humanity, studies conducted on people belonging to this very specific demographic have served as the foundation for a majority of modern scientific efforts. Although this trend has been acknowledged, challenged, and serious revision efforts have been undertaken, the fact remains that many participants in experimental studies are still WEIRD (Western Educated Industrialized Rich Democratic) (Henrich, Heine, and Norenzayan 2010) adults, usually of university age. Within human subjects research, children and non-WEIRD participants are typically recruited only when a research question specifically necessitates it. The convenience of WEIRD adult sampling is undeniable; however, relying too heavily on a single demographic creates the same sorts of gaps in current literature previously caused by recruiting only men – researchers fall victim to the seductive notion that their research model is generalizable beyond its reasonable bounds.

This error, while acknowledged in conference questions, personal communication, and the occasional – and very important – publication (Shea 2006; Finlay 1997; Lillehammer 1989; Högberg 2008; Neubauer 2018), continues to befall the sampling practices of experimental archaeological work. The vast majority of experimental work in stone toolmaking still utilizes an adult-only sample, although the sex distribution of novice participants, in particular, has somewhat equalized, or, in some cases (e.g. Cataldo, Migliano, and Vinicius 2018), has shifted toward a female bias. Typically, members of an expert sample are majority male, though this trend may be an artifact of the historically lower numbers of women academic archaeologists in professorial positions with research labs. In some published papers, the actual age distribution of participants is not disclosed, except to say that the participants are, indeed, adults, and little to nothing is mentioned about other demographic information, such as gender identity, socioeconomic status, race/ethnicity, academic performance/highest level of education completed, etc. Although not all of these identifiers may be relevant or even discernable in archaeological lithic assemblages, and therefore may not be pertinent to evolutionary questions, such details would be useful in better understanding how demographic circumstances affect modern human knappers. The present study focuses on the developmental differences in stone knapping skill acquisition and does not attempt to address all aspects of participant identity, as listed above. With that said, while many researchers may privately acknowledge that the absence of children in experimental archaeology is problematic, little has been done to rectify the issue.

To date, experimental lithic work conducted with modern children is still rare, and what little has been done has been published as book chapters and in lesser-known scientific journals, if it is written for publication, at all. This publishing trend emphasizes that although there is general interest in the theoretical aspects of the archaeology of childhood, there is still much work that needs to be done to elevate the profile of this important topic, and this may be accomplished by establishing a firm empirical basis for the study of ancient children and their role in material culture production and evolution. Researchers Sternke and Sørensen (2005) lay some of this foundation through what may be the first stone knapping experiment conducted with child participants. In their study, the authors recruited a total of eight subjects, six of whom were children (ages 6 – 11 years, mean age: 8.92 years) and two of whom were adults (both aged 25 years). The amount of knapping experience for these individuals ranged between 1h 45 mins and 2h 50 mins of practice time for the children and between 1 month and 2 years of practice for the adults. Participants were asked to produce a range of Later Mesolithic (Ertebølle) tools and received verbal instruction from the authors on how to do so. The authors mention, but do not discuss, their observations regarding the individual differences in motor control, mental capability, and hand-eye coordination between subjects of different ages. Although the novice knappers made mistakes typical of beginners (e.g. hinged flake terminations, lack of standardization in final shape), all participants were able to produce simple flakes and scrapers, though the children could not produce the more complicated form of the transverse arrowhead. Sternke and Sørensen conclude that a cognitive understanding of the steps involved in a reduction sequence, more than any physical limitations imposed by knapping, are what may drive the differences between child and adult stone knapping, but more importantly, they suggest that such differences do exist, though conclusions are limited by a very small sample size. Sternke and Sørensen provide a promising beginning for work done with child knapping, but it is clear that further research is needed to hone in on differences between child and adult knappers and what motivates them.

Identifying Novice Knappers in Archaeological Assemblages

Identifying individuals in archaeological lithic assemblages is a notoriously difficult task, if it is even possible with our current methods and technology (Eren, Bradley, and Sampson 2011). Because lithics are often found without direct contextualizing information about their creators (e.g. fossilized remains), assemblages have historically been analyzed at the group level, with the acknowledgement that a group of individual agents – unknowable in identity or number – would have been responsible for their manufacture. Despite this difficulty, however, there has been a recent call to action advocating the role of the individual in archaeological assemblages, generally (Gamble and Porr 2005), and in the production of stone tools, specifically (Hopkinson and White 2005; Gowlett 2005; Pope and Roberts 2005). Although it is true that evolution does not act on the scale of the individual and that to address evolutionary questions, group level analyses are necessary, identifying individual knappers in the lithic record remains an important point of inquiry to address questions such as those relating to the evolution of teaching. For example, because of developmental differences in cognition and motor control, which would be present in the ontogeny of all hominin species, expert knappers may approach the tutelage of their novices differently based on whether the inexperienced knappers were children or adults. Because of this, being able to determine the age of individual stone toolmakers in the archaeological record may elucidate not just the identity of the novice knappers, but also how skilled knowledge of stone toolmaking was passed between individuals within a knapping community.

Novices, often defined in this literature in general terms as individuals who have not yet gained expertise, are associated with suboptimal knapping techniques, such as premature abandonment of cores (Castañeda 2018), improper flake removal (Dugstad 2010; Goldstein 2018; Karlin and Julien 2019), relative lack of complexity in reduction sequences (Maloney 2019), miniaturization (Knight 2017), and overall irregularity of shape in finished forms (Johansen and Stapert 2012; Neubauer 2018). However, for the most part, criteria used to identify novices reflect the intuitions of lithic analysts rather than empirically demonstrated patterns. Additionally, it is important to note that the majority of work done on identifying novices has involved assemblages found in European sites (Castañeda 2018; Dugstad 2010; Högberg 2008; Johansen and Stapert 2012; Karlin and Julien 2019), with some notable exceptions (Goldstein 2018; Cunnar 2015; Knight 2017; Maloney 2019; Neubauer 2018), and in relatively recent time periods. Although this research sets a foundational methodological precedent, these efforts should be applied to a more diverse range of site locations, time periods, and lithic industries in order to establish a more complete representation of novice knappers' work across space and time.

For decades (Lillehammer 2015), researchers have made efforts to address the opacity of children's influence in the archaeological record, with emphasis on identifying novice work at lithic production sites. Despite the undoubtable overlap between novice knappers and child knappers, it should also not be assumed that these groups are synonymous (Ferguson 2008; Högberg 2018). Although it is likely that ancient hominins would have begun the process of learning stone knapping during childhood (Shea 2006), the extent to which children would have engaged with raw materials and production is less certain, for reasons including the rarity and quality of raw materials, safety precautions, and the physical and cognitive affordances of knapping (Ferguson 2008; Kamp 2015). More probable is the scenario wherein knapping skills are acquired through a gradual scaffolding process that begins with passive observation of knapping activities and graduates into direct handling of raw materials and attempts to produce lithic products. This process would have begun early in childhood and continued through adolescence and into early adulthood (Nowell and French 2020; Riede et al. 2018), which brings us back to the following question: how old were the novice knappers identified in lithic assemblages? This inquiry necessitates a consideration of how childhood is defined, both across evolutionary time and existing cultural contexts (Högberg 2008; Kamp 2015), and highlights the importance of a research program that investigates knapping skill acquisition across development.

SKILL ACQUISITION AND STONE TOOLMAKING

Studies of skill acquisition and the development of expertise, although relatively new in the stone toolmaking literature, have a long history in the psychological sciences. Dating back over a century (Bryan and Harter 1899), research on skilled behavior emphasizes the role of regular, deliberate practice in the acquisition of expertise (Ericsson, Krampe, and Tesch-Romer 1993) and has demonstrated a phenomenon known as the 'power-law of practice' (Newell and Rosenbloom 1981), which describes an initial period of rapid increase in proficiency followed by diminishing returns as an individual's skill level flattens to a local optimum. This curve is further characterized by more subtle alternating periods of improvement and plateau as learners engage in cognitive 'chunking' of acquired information (Guida et al. 2012). Through chunking, individuals reduce strain on their working memory capacity by storing summary chunks of information in their long-term memory, which aids in quicker retrieval and, consequently, in better task performance. The relationship between chunking and expertise is well-documented, and studies of skill acquisition in stone toolmaking have also recently demonstrated this trend (Geribàs, Mosquera, and Vergès 2010; Pargeter, Khreisheh, and Stout 2019).

For example, Geribàs and colleages (2010) reported that novice knappers are more likely to approach stone knapping as a sequence of disjointed events, whereas expert knappers perceived knapping as clusters of related behaviors. This difference in perspective was evident in the behavior sequences recorded for expert and novice knappers. Whereas the experts were likely to repeat patterned clusters of behaviors, novices demonstrated no patterning in their behavioral output. Similarly, expert knappers were more likely to use a broader range of the behaviors observed by the study's ethogram (e.g. clustered on factors such as shaping of the core, frequency of rotating the core versus percussion, and knapping posture), and novice knappers were more prone to be retricted in their stone knapping behavior. In fact, novices rarely used rotation during their knapping, instead relying almost soley on percussion. Together, these observations indicate that novice knappers adopted a strategy of using a small range of behaviors, perhaps those that they felt most comfortable using or that they felt yeilded the most obvious results, repeatedly and without direction, while expert knappers were stragegic in their approach to bifacial reduction.

In addition to chunking-related behaviors, expert performance across a variety of activities, including in infants learning to use a baby bouncer (Goldfield, Kay, and Warren 1993) and stone knappers regulating the velocity and trajectory length of their hammerstones according to hammerstone weight (Bril et al. 2010), appears to minimize unnecessary movements and inefficient energy expenditure, thus optimizing the relative effort required for optimum output. This ability is acquired through flexible management of environmental constraints and adapting to unfamiliar task parameters, such as in the case of stone bead knapping in India (Roux, Bril, and Dietrich 1995). In their study of stone bead knappers, Roux, Bril, and Dietrich (1995) also noted that craftsmen who underwent a longer apprenticeship period (seven to ten years) were more adept at organizing what the authors refer to as the elementary movements (i.e. the most basic unit of an action sequence) into sub-goals during a planned knapping reduction sequence than were the craftsmen who received less training (two to three years). Most notably, the less experienced craftsmen spent less time, or skipped entirely, calibrating their beads and reducing the beads' stone crests during shaping, whereas the more experienced craftsmen routinely followed this procedure, and, as a result, produced more uniform bead shapes across participants. Stone knapping, at its most basic, is a craft that applies the mechanics of conchoidal fracture to produce the controlled removal of flakes from a core. Although the particular physical principles of this type of fracture are not fully understood, stone knappers must develop an intuitive understanding of conchoidal fracture in order to employ it effectively during stone toolmaking.

One of the primary limitations of experimental stone knapping studies is the difficulty in recruiting participants for adequate training periods. Because expertise in stone toolmaking is acknowledged, through ethnographic accounts (Stout 2002, Roux, Bril, and Dietrich 1995), to take years to acquire, the most ecologically valid studies of knapping skill acquisition would require a longitudinal design; however, the demands of such a design, both in participant time and expense of raw stone materials, makes this effort prohibitive. Relatively short training times are ubiquitous in the experimental stone knapping literature, with some spanning only five minutes (Morgan et al. 2015), but despite the substantial apprenticeship period required to gain true mastery of stone toolmaking, participants have shown improvement in their knapping skills and knowledge within the span of the comparatively much shorter experimental paradigm. For example, Pargeter, Khreisheh, and Stout (2019) asked participants to commit upwards of 90 hours of knapping practice over the duration of the study, with assessments spaced at 10-hour intervals. Knapping skill was initially recorded by the instructor at the end of each session using a 10-point scale, which assigned participants scores that were used to track learning progress. These scores were later compared with scores produced by a random forest regression model developed from the measurement data of the experimental handaxes produced during the assessments. The model employed many of the same criteria of stone knapping skill as the subjective scoring (e.g. handaxe thinning and shaping, use of bifacial reduction strategy), but instead looked at nine measurement variables taken from the assessment handaxes. From these data emerged a distinct power-law curve of participant skill increase, with rapid improvement taking place within the first 30 hours of practice. Individual participant success during this early stage did correlate with high performance in psychometric tasks of planning and problem solving (Tower of London) and ‘set shifting,’ or the ability to demonstrate flexibility amid changing rules (Wisconsin Card Sort). Past this initial stage, however, practice density was a greater predictor of skill increase.

Thus, in order to gain expertise in stone toolmaking, novice knappers must master the following: the chunking of elementary actions into systematic sub-goals, optimization of energy expenditure and minimization of unnecessary movements, the ability to adapt flexibly to the knapping environment (e.g. raw material), an understanding of conchoidal fracture mechanics, and an ability to identify salient aspects of knapping affordances (e.g. posture and core handling, viable core morphology, appropriate hammerstone selection, and correct striking angle and velocity). To acquire these skills, they need to engage in deliberate, individual practice, usually over a lengthy period of upwards of ten years (Ericsson, Krampe, and Tesch-Romer 1993; Ericsson 2008). Evidence of their acquired mastery is demonstrated in their lithic output, namely in factors such as flake size (i.e. larger flakes are indicative of more skilled knapping) and utility (i.e. flakes with more cutting edge are considered to be more ‘useful’).

CHILDHOOD: EVOLUTION AND DEVELOPMENT

Hominin Life History and the Evolution of Childhood

Hominin life history is marked by several notable differences from the life history traits of other primate species, namely cooperative breeding (Hrdy 2009), shorter interbirth intervals (Humphrey 2010; Nakahashi, Horiuchi, and Ihara 2018), longer lifespans (Neill 2014), and an extended juvenile period (Konner 2010; Walker et al. 2006). Though hominin life history may not be as unique as it is sometimes portrayed (Miller, Churchill, and Nunn 2019), an extended juvenile period, or childhood, is a prominent human characteristic, one that is rare in other species and never so exaggerated. The evolutionary origins of childhood as a distinct hominin life stage are debated both in timing and cause. Although there is evidence to suggest that childhood could have emerged as a distinct life history stage as early as *Homo habilis* (~2.4 – 1.4

million years ago) (Bogin 1997; Bogin and Varea 2020), others contend, citing evidence of the timing of the M1 molar eruption in ancient *Homo erectus* juvenile specimens, that this date is much too early and place the emergence of childhood with archaic modern humans, *Homo antecessor* and *Homo heidelbergensis* (Thompson and Nelson 2011). The reason behind the expansion of the juvenile period is also under dispute, with some arguing that it is a byproduct of cooperative breeding and earlier age of weaning (Bogin 1997) and others asserting its adaptiveness as a time of learning social norms and skilled behaviors (Street et al. 2017). In modern humans, it is apparent that childhood is a time of rapid development across domains, such as cognitive ability and motor skill, which has been well-documented by decades of work in developmental psychology. Though the stages of childhood have been described in various ways by prominent figures in the field (Piaget 1952; Vygotsky 1978), childhood can be broadly characterized in three main stages: early (birth to 6 years old), middle, (7 to 12 years old), and adolescence (13 to 18 years old). Although each developmental period is host to its own array of cognitive, motor, and psychosocial changes, this section will focus on those that take place during middle childhood, as it is demonstrated to be the age during which skilled learning begins in modern forager societies (Crittenden 2016; Boyette and Hewlett 2017), which serve as the closest living analog to the ecological conditions of our hominin ancestors.

Motor Development

A unique feature of human life history (DelGiudice 2018), middle childhood is marked by the rapid maturation of many of the motor, cognitive, and psychosocial processes and abilities that are hypothesized to be important to stone toolmaking ability. The transition to middle childhood is marked by an increase in aptitude for both gross and fine motor skills (DelGiudice, 2018), for mental imagery tasks associated with motor control (Gabbard 2009; Spruijt, van der Kamp, and Steenbergen 2015), and for motor coordination in joint attention tasks (Satta et al. 2017).

Much work in this first domain has been conducted on the development of children's gait and posture, with and without the added difficulty of a dual constraint task (Al-Yahya et al. 2011; Chauvel et al. 2017; Fabri et al. 2017; Gill, Yang, and Hung 2017; Saxena et al. 2017; Schaefer et al. 2015). For example, Gill, Yang, and Hung (2017) performed a cross-sectional study of children in three age groups - young (4 - 6 years old), middle (7 - 9 years old), and old (10 - 13 years old). The children in each group completed three tasks: 1) finger rotation, 2) obstacle crossing, and 3) carrying a box while walking. In each of these tasks, the children in the young group were more variable in their motor capabilities than those in either of the older groups. The authors indicated these findings as evidence for continued motor maturation during early into middle childhood, which is consistent with the general trend of developmental shifts in these stages of childhood.

Additionally, children improve not just in their ability to perform motor tasks physically, but also in their competence in tasks of motor imagery (Gabbard 2009; Spruijt, van der Kamp, and Steenbergen 2015), which ask participants to imagine themselves engaging in different sorts of activities. This research paradigm is based on the assertion that these motor imagery tasks will provide a 'window' into the mental processes of motor actions. Gabbard (2009), in a review of the motor imagery literature, mentions that the most prominent differences between adult and child competence at motor imagery tasks were found in variables related to bodily awareness. More specifically, children and adolescents appeared to have less developed internal models, perhaps relating to the still-developing parts of their neural circuitry, such as that found in the parietal cortex. Interestingly, Gabbard also notes that children under the age of 7 do not seem to possess the ability to form motor images of themselves and that this capacity emerges around age 7 year. Such data dovetail the robust body of evidence on the developmental changes that take place during the 5 to 7 shift, both in cognitive and motor domains (Weisner 1996).

Finally, with the emergence of theory of mind, discussed below, children are able to engage in tasks involving joint attention (Satta et al. 2017). Satta et al. (2017) investigated the age at which this cognitive-motor function emerges with a sample of children, ages 6 to 9, and a sample of adult participants. Their experiment consisted of one task, using an isometric joystick to direct a visual cursor from a central to a peripheral target, across two conditions, one where the participant acted in isolation and another where two participants coordinated on their task. Satta et al. (2017) concluded that although children ages 6 and 7 were able to coordinate the onset of their joint movement task, they were unable to achieve the level of synchrony to succeed fully at the task, which was to act in unison from the beginning of the cursor movement to the end. The 8-year-old participants seemed to be more apt at their solo performance, and the researchers note that past this age, children were more likely to attend to their peer's performance in an effort to achieve behavioral synchrony through joint attention. These results indicate that the motor and cognitive skills required to be an active participant in advanced forms of social learning, such direct-active teaching, undergo development during middle childhood, given that a child's aptitude in these domains changes over the course of this developmental stage.

Cognitive Development and Theory of Mind

Cognitive development in middle childhood is dominated by the maturation of executive functioning abilities and the stable emergence of theory of mind, a trait hypothesized to play a critical role in the social transmission of stone toolmaking (Stout 2011). Executive functioning is often defined as an umbrella term for the network of processes involved in intentional, goal-directed behavior (Anderson 2002). It specifically includes inhibition, or the ability to self-regulate behavioral responses; cognitive flexibility, which is the ability to switch fluidly between tasks and to quickly adapt to new or changing information; and working memory, which allows an individual to retain and manipulate information for a short period of time (Cantin et al., 2016). These individual cognitive capacities, respectively, enable an individual to delay immediate gratification in favor of a more distant but more desirable outcome, to incorporate external feedback into an internal mental schema, and to remember a sequence of events for a long enough period of time to enact experience-based behavioral change, skills that are hypothesized to be relevant to stone toolmaking acquisition (Stout and Khreisheh 2015; Pargeter, Khreisheh, and Stout 2019).

Although not entirely absent in earlier developmental stages, during middle childhood, children achieve proficiency in the above abilities. For example, in a cross-sectional study involving 7- to 10-year-olds, Cantin et al. (2016) examined the effects of age on success in tasks testing the components of executive functioning, which included a Digit Span subtest for working memory, a Color Work Stroop task for inhibition, and a modified Dimensional Change Card Sorting task for flexibility. Though Cantin and colleagues were interested in the degree to which the relationship between age and executive functioning could predict academic outcomes, such as reading comprehension and mathematics performance, and social understanding by way of theory of mind, their study provides a clear example of age-related developmental changes in executive function abilities. More specifically, these authors applied a model-based approach to analyzing their data. Their initial model treated the components of executive function as parts of a unified whole, and therefore did not examine the individual tri-part relationships between each category of variables (e.g. age, working memory, and reading comprehension; or age, inhibition, and theory of mind). When this model did not reveal any significant results, the authors built a model that analyzed working memory, inhibition, and flexibility separately, and with this change, patterns emerged in the data to corroborate the relationship between age and aptitude at tests of executive function.

The results of this study support two important trends in developmental literature. First, that age does correlate with cognitive development. Second, that the components of executive functioning operate, and likely develop, independently from one another. Cantin et al. (2016) suggest that it is during middle childhood that working memory, inhibition, and cognitive flexibility appear to integrate into what is known collectively as executive function. This account of cognitive development is consistent with a modular conception of the brain and its systems, given that several regions of the brain act together to perform a generalized task (Sporns and Betzel 2016).

During middle childhood, children also grow in their capacity for theory of mind. This cognitive trait, defined as an individual's ability to acknowledge that others experience inner thought lives similar to but different than one's own, emerges in early childhood reliably around the age of 4 years (Flavell 1999; Lagattuta et al. 2015). Some argue that infants as young as 14-months-old are able to demonstrate false belief understanding, as measured through gaze time during a false belief task (De Bruin and Newen 2012). A standard and well-established measurement for detecting theory of mind, in children and adults alike, the false belief task typically involves a narrative like the one following: Individual A hides an object in place A. When Individual A leaves the area, and unknown Individual B enters, moves the hidden object to place B, and then leaves. Individual A returns, looks for the hidden object, and the narrative ends. After watching a sequence like this, children participating in a false belief task will be asked where Individual A will look for the hidden object - in place A, where they left it, or in place B. Children pass the task if they are able to articulate to an experimenter that Individual A will look for the hidden object in place A because they are not aware that Individual B moved the object to place B. Children who have not yet developed theory of mind will answer that Individual A will look for the object in place B, because they do not yet understand that Individual A has a different set of knowledge than their own.

Piaget and Inhelder (1956) also document the developmental shift from what they call egocentrism to the emergence of theory of mind in their Three Mountain Problem, which tests children on whether they can imagine another person's point of view. Like the false belief task described above, children in this experimental paradigm are asked to describe what a doll can see from different vantage points on the mountain. Children are allowed to inspect the mountain from all angles and therefore are familiar with all of its features. If they are able to articulate that the doll 'sees' different aspects of the mountain when viewing it from a perspective different from their own, children are said to have developed past the egocentric phase of early childhood, a transition that occurs around age 7 year and corresponds with entry into middle childhood (Piaget and Inhelder 1956).

Advancement in theory of mind capabilities does not cease with its initial emergence in this transition from early to middle childhood. As a part of this developmental shift, a typically developing child will also learn to make inferences about another's mental state, such as their experience of specific emotions or the recognition of shared knowledge (e.g. a typical example of second order intentionality: 'she knows that I know about her vacation to the beach') (Flavell

1999). This more advanced version of theory of mind is critical for the development of empathy, which itself is fundamental to an individual's ability to learn socially.

Wang et al. (2016) investigated the continued development of theory of mind in children ages 8- and 10-years-old. They tested both for egocentrism, defined as an incorrect choice on a perspective-taking task, and for the effects of language complexity on the resurfacing of egocentric behavior in children who exhibited false belief understanding in the initial task. This latter condition was designed to examine the role of language comprehension in developing theory of mind. The authors found that, generally, the 8-year-old participants were more likely than their 10-year-old counterparts to respond egocentrically to the baseline condition false belief task. However, the performance in both age groups suffered in the condition where their instructions for the second task were given to them in more complex language. Here, complexity was operationalized as increasing degrees of specificity. For example, in the complex condition, participants received instructions resembling the phrase, 'nudge the small ball up one slot,' whereas those in the simple condition were just told to 'nudge the small ball.' Here, these authors demonstrate a relationship between language development and an increase in aptitude for theory of mind tasks. This connection should not be understated, as language and cognition inextricably linked in human mental processes.

Taken together, the maturation of executive functioning and theory of mind during middle childhood provides a cognitive scaffold that enables individuals to engage in more complicated, skilled behaviors that may be transmitted through social learning. These developmental changes coincide with a behavioral trend felt globally by individuals of this age, namely that children between the ages of 7 and 12 years are expected to begin participating in tasks that more closely resemble adults' work (Weisner 1984).

STONE KNAPPING: COMPARING ADULTS AND CHILDREN, EXPECTED OUTCOMES

Given the differences of motor and cognitive ability between children and adults and with evidence of differential performance in skill-based activities such as foraging, it stands to reason that children and adults will also perform differently in a task of stone toolmaking. Though the amount of force and finesse required to detach a flake varies depending upon the type of material a knapper uses (Eren et al. 2014; Key, Proffitt, and de la Torre 2020; Li et al. 2022), certain motor ability prerequisites still remain for an individual to be able to produce flakes of a usable size. These include upper body strength, motor accuracy, and manual coordination. Hand grip strength is often used as a proxy for overall upper body strength, and it is no surprise that average grip strength varies dramatically across the lifespan and between males and females (Bardo et al. 2021). Children up through age 10 years resemble each other in grip strength output, regardless of sex, but upon entering the later stages of middle childhood and into adolescence, males begin to outstrip their female peers (Häger-Ross and Rösblad 2002). By the age of 18 years, an individual's grip strength reaches its adult capacity, which peaks at 36 years old, notwithstanding any deliberate weight training exercises that would modify these upper limits, after which grip strength steadily declines with senescence (Nahhas et al. 2010). Because stone toolmaking does require some measure of upper body strength, regardless of the material used, one would expect there to be a lower limit of how strong an individual must be in order to successfully detach stone flakes. It is likely that this lower limit is far below the average adult's grip strength and therefore has gone undetected as nearly all stone knapping experiments have been conducted with adult participants. With children, however, it is possible that their relatively lower grip strengths will limit not just the size of flake they are able to detach, but also how many flakes they are able to produce from a core.

As with grip strength and overall fitness, adults tend to outperform children in psychometric tests of inhibition control (Leon-Carrion, Garcia-Orza, and Perez-Santamaria 2004), motor accuracy (Smits-Engelsman, Sugden, and Duysens 2006), and mental manipulation of objects (Childs and Polich 1979). Although children experience rapid increases in these abilities from early to middle childhood, their progress often slows, or even regresses, before reaching adult levels (Best and Miller 2010). In a motor learning task, wherein adults and children were asked to perform a discrete, coordinated arm movement, children required longer periods of practice and more consistent feedback to gain proficiency (Sullivan, Kantak, and Burtner 2008). The importance of practice density in learning motor tasks was also observed in a Frisbee or ropeball throwing task, where both adults and children benefited from blocked training periods over sporadic practice (Zipp and Gentile 2010). In all, although both adults and children require practice to successfully acquire new motor skills, one would expect children to take more time and need more practice and encouragement in order to gain proficiency. As this applies to stone toolmaking, children with higher psychometric scores (i.e. those approaching adult levels) may more quickly acquire and internalize the principles of knapping that employ cognitive resources, such as being deliberate in selecting and striking core platforms (i.e. inhibition control), identifying and remembering salient aspects of core morphology (i.e. mental rotation), and striking the core with accuracy (i.e. motor accuracy).

Although children are generally disadvantaged, compared to adults, in the motor and cognitive abilities hypothesized to contribute to stone toolmaking success, the difference in performance between adults and children in a skill-based task may have less to do with their capacity to understand the task and more to do with the different trade-offs they face due to bodily affordances. In the case of children's coastal foraging among the Meriam of the Torres Strait Islands, children as young as 5 or 6 years can reliably identify and collect appropriate prey, with modest guidance from adults (Bird and Bliege Bird 2002). What limits their foraging output is not their knowledge of which prey items are of high-value – instead, their prey selection is more restricted than adults' due to the strength required to process some prey items and their relatively slower walking speeds.

In summary, based on the cognitive and motor differences between children and adults, listed above, we would expect middle childhood aged children to produce fewer stone flakes that are smaller and less refined than those produced by adolescents and adults up through middle age. Which features factor most prominently in this difference remains a point of investigation, as does the role of individual differences between knappers, irrespective of age. The following study seeks to contribute to addressing the questions of developmental influences and individual differences in stone toolmaking in an experimental setting with modern human participants.

WORKS CITED

Al-Yahya, E., H. Dawes, L. Smith, A. Dennis, K. Howells, and J. Cockburn. 2011. Cognitive motor interference while walking: a systematic review and meta-analysis. *Neurosci Biobehav Rev* 35 (3):715-28.

Ambrose, Stanley. 2001. Paleolithic Technology and Human Evolution. *Science* 291:1748-1753.

Anderson, P. 2002. Assessment and development of executive function (EF) during childhood. *Child Neuropsychol* 8 (2):71-82.

Antón, Susan C., and J. Snodgrass. 2012. Origins and Evolution of Genus *Homo*. *Current Anthropology* 53 (S6):S479-S496.

Ascher, Robert. 1961. Analogy in Archaeological Interpretation. *Southwestern Journal of Anthropology* 17:317-325.

Bardo, Ameline, Tracy L. Kivell, Katie Town, Georgina Donati, Haiko Ballieux, Cosmin Stamate, Trudi Edginton, and Gillian S. Forrester. 2021. Get a Grip: Variation in Human Hand Grip Strength and Implications for Human Evolution. *Symmetry* 13 (7).

Barsalou, L. W. 2008. Grounded cognition. *Annu Rev Psychol* 59:617-45.

Baxter, Jane Eva. 2008. The Archaeology of Childhood. *Annual Review of Anthropology* 37 (1):159-175.

Best, J. R., and P. H. Miller. 2010. A developmental perspective on executive function. *Child Dev* 81 (6):1641-60.

Bird, Douglas W., and Rebecca Bliege Bird. 2000. The Ethnoarchaeology of Juvenile Foragers: Shellfishing Strategies among Meriam Children. *Journal of Anthropological Archaeology* 19 (4):461-476.

Repeated Author. 2002. Children on the Reef: Slow Learning or Strategic Foraging? *Human Nature* 13:269-297.

Blasi, D. E., S. Wichmann, H. Hammarstrom, P. F. Stadler, and M. H. Christiansen. 2016. Sound-meaning association biases evidenced across thousands of languages. *Proc Natl Acad Sci U S A* 113 (39):10818-23.

Bogin, B., and C. Varea. 2020. Evolution of Human Life History. In *Evolutionary Neuroscience*.

Bogin, Barry. 1997. Evolutionary Hypotheses for Human Childhood. *Yearbook of Physical Anthropology* 40:63-89.

Bordes, François. 1961. Mousterian Cultures in France. *Science* 134:803-810.

Boroditsky, L. 2001. Does language shape thought? Mandarin and English speakers' conceptions of time. *Cogn Psychol* 43 (1):1-22.

Boyette, A. H. 2013. Social Learning During Middle Childhood Among Aka Foragers and Ngandu Farmers of the Central African Republic, Anthropology, Washington State University.

Boyette, A. H., and B. S. Hewlett. 2017. Autonomy, Equality, and Teaching among Aka Foragers and Ngandu Farmers of the Congo Basin. *Hum Nat* 28 (3):289-322.

Boyette, A. H., and Barry S. Hewlett. 2017. Teaching in Hunter-Gatherers. *Review of Philosophy and Psychology* 9 (4):771-797.

Bril, B., R. Rein, T. Nonaka, F. Wenban-Smith, and G. Dietrich. 2010. The role of expertise in tool use: skill differences in functional action adaptations to task constraints. *J Exp Psychol Hum Percept Perform* 36 (4):825-39.

Bryan, William Lowe, and Noble Harter. 1899. Studies on the Telegraphic Language: The Acquisition of a Hierarchy of Habits. *Psychological Review* 6:345-375.

Burkart, J. M., S. B. Hrdy, and C. P. Van Schaik. 2009. Cooperative breeding and human cognitive evolution. *Evolutionary Anthropology: Issues, News, and Reviews* 18 (5):175-186.

Burkart, J. M., M. N. Schubiger, and C. P. van Schaik. 2017. The evolution of general intelligence. *Behav Brain Sci* 40:e195.

Byrge, L., O. Sporns, and L. B. Smith. 2014. Developmental process emerges from extended brain-body-behavior networks. *Trends in Cognitive Sciences* 18 (8):395-403.

Byrne, Richard W. 2000. Evolution of Primate Cognition. *Cognitive Science* 24 (3):543-570.

Byrne, Richard W., and Lisa G. Rapaport. 2011. What are we learning from teaching? *Animal Behaviour* 82 (5):1207-1211.

Cantin, R. H., E. K. Gnaedinger, K. C. Gallaway, M. S. Hesson-McInnis, and A. M. Hund. 2016. Executive functioning predicts reading, mathematics, and theory of mind during the elementary years. *J Exp Child Psychol* 146:66-78.

Caro, T.M., and M.D. Hauser. 1992. Is There Teaching in Nonhuman Animals? *The Quarterly Review of Biology* 67:151-174.

Castañeda, Nuria. 2018. Apprenticeship in Early Neolithic Societies: The Transmission of Technological Knowledge at the Flint Mine of Casa Montero (Madrid, Spain), ca. 5300–5200 cal BC. *Current Anthropology* 59 (6):716-740.

Cataldo, D. M., A. B. Migliano, and L. Vinicius. 2018. Speech, stone tool-making and the evolution of language. *PLoS One* 13 (1):e0191071.

Chauvel, G., E. Palluel, A. Brandao, G. Barbieri, V. Nougier, and I. Olivier. 2017. Attentional load of walking in children aged 7-12 and in adults. *Gait Posture* 56:95-99.

Childs, Michael K., and John M. Polich. 1979. Developmental Differences in Mental Rotation. *Journal of Experimental Child Psychology* 27:339-351.

Clark, A. 2006. Language, embodiment, and the cognitive niche. *Trends Cogn Sci* 10 (8):370-4.

Clark, Andy. 1999. An embodied cognitive science? *Trends in Cognitive Sciences* 3:345-351.

Coolidge, Frederick L., and Thomas Wynn. 2016. An Introduction to Cognitive Archaeology. *Current Directions in Psychological Science* 25 (6):386-392.

Crabtree, Don E., and E.L. David. 1968. Experimental Manufacture of Wooden Implements with Tools of Flaked Stone. *Science* 159:426-428.

Crittenden, A. N. 2016. Children's foraging and play among the Hadza. In *Childhood: Origins, Evolution, and Implications*, edited by C. L. Meehan and A. N. Crittenden: University of New Mexico Press.

Cunnar, Geoffrey E. 2015. Discovering Latent Children in the Archaeological Record of the Great Basin. *Childhood in the Past* 8 (2):133-148.

Currie, Adrian, and Anton Killin. 2019. From things to thinking: Cognitive archaeology. *Mind & Language* 34 (2):263-279.

Dart, Raymond A. 1925. Australopithecus africanus: The Man-Ape of South Africa. *Nature* 115 (2884):195-199.

De Bruin, L. C., and A. Newen. 2012. An association account of false belief understanding. *Cognition* 123 (2):240-59.

de la Torre, I. 2011. The origins of stone tool technology in Africa: a historical perspective. *Philos Trans R Soc Lond B Biol Sci* 366 (1567):1028-37.

Repeated Author. 2016. The origins of the Acheulean: past and present perspectives on a major transition in human evolution. *Philos Trans R Soc Lond B Biol Sci* 371 (1698).

de la Torre, I., and R. Mora. 2009. Remarks on the Current Theoretical and Methodological Approaches to the Study of Early Technological Strategies in Eastern Africa. In *Interdisciplinary Approaches to the Oldowan*, edited by E. Hovers and D. R. Braun: Springer.

Repeated Author. 2018. Oldowan technological behaviour at HWK EE (Olduvai Gorge, Tanzania). *J Hum Evol* 120:236-273.

de Sassure, Ferdinand. 2013. Nature of the Linguistic Sign and Synchronic and Diachronic Law. In *Readings for a History of Anthropological Theory*, edited by P. A. Erickson and L. D. Murphy: University of Toronto Press.

Deacon, T.W. 1997. *The symbolic species: The coevolution of language and the brain*. New York: Norton.

Dehaene, Stanislas. 2011. *The Number Sense: How the Mind Creates Mathematics*. Second ed: Oxford University Press.

DelGiudice, M. 2018. Middle Childhood: An Evolutionary-Developmental Synthesis. In *Handbook of Life Course Health Development*, edited by N. Halfon, C. B. Forrest, R. M. Lerner and E. M. Faustman. Cham (CH).

Depew, David J. 2003. Baldwin and His Many Effects. In *Evolution and Learning: The Baldwin Effect Reconsidered*, edited by B. H. Weber and D. J. Depew. Cambridge, Massachusetts: The MIT Press.

Dibble, Harold L., and John C. Whittaker. 1981. New Experimental Evidence on the Relation Between Percussion Flaking and Flake Variation. *Journal of Archaeological Science* 8:283-296.

Dingemanse, M., D. E. Blasi, G. Lupyan, M. H. Christiansen, and P. Monaghan. 2015. Arbitrariness, Iconicity, and Systematicity in Language. *Trends Cogn Sci* 19 (10):603-615.

Dissanayake, Ellen. 2017. Ethology, Interpersonal Neurobiology, and Play. *American Journal of Play* 9:143-168.

Du, Y., N. C. Valentini, M. J. Kim, J. Whitall, and J. E. Clark. 2017. Children and Adults Both Learn Motor Sequences Quickly, But Do So Differently. *Front Psychol* 8:158.

Dugstad, Sigrid A. 2010. Early child caught knapping. A novice Early Mesolithic flintknapper in south-western Norway. *Socialisation: Recent Research on Childhood and Children in the Past. Stavanger: AmS-Skrifter* 23:65-74.

Dunbar, Robin I.M. 1998. The Social Brain Hypothesis. *Evolutionary Anthropology*:178-190.

Eren, Metin I., B. A. Bradley, and C. Garth Sampson. 2011. Middle Paleolithic Skill Level and the Individual Knapper: An Experiment. *American Antiquity* 76:229-251.

Eren, Metin I., Stephen J. Lycett, Robert J. Patten, Briggs Buchanan, Justin Pargeter, and Michael J. O'Brien. 2016. Test, Model, and Method Validation: The Role of Experimental Stone Artifact Replication in Hypothesis-driven Archaeology. *Ethnoarchaeology* 8 (2):103-136.

Eren, Metin I., Christopher I. Roos, Brett A. Story, Noreen von Cramon-Taubadel, and Stephen J. Lycett. 2014. The role of raw material differences in stone tool shape variation: an experimental assessment. *Journal of Archaeological Science* 49:472-487.

Ericsson, K. A. 2008. Deliberate practice and acquisition of expert performance: a general overview. *Acad Emerg Med* 15 (11):988-94.

Ericsson, K. Anders, Ralf Th. Krampe, and Clemens Tesch-Romer. 1993. The Role of Deliberate Practice in the Acquisition of Expert Performance. *Psychological Review* 100:363-406.

Fabri, T. L., K. E. Wilson, N. Holland, A. Hickling, J. Murphy, P. Fait, and N. Reed. 2017. Using a dual-task protocol to investigate motor and cognitive performance in healthy children and youth. *Gait Posture* 54:154-159.

Falk, D. 2009. *Finding out tongues: Mothers, infants, and the origins of language*: Basic Books.

Repeated Author. 2012. Hominin paleoneurology: where are we now? *Prog Brain Res* 195:255-72.

Ferguson, Jeffrey R. 2008. The When, Where, and How of Novices in Craft Production. *Journal of Archaeological Method and Theory* 15 (1):51-67.

Finlay, Nyree. 1997. Kid knapping: The missing children in lithic analysis. In *Invisible people and processes: Writing gender and childhood into European archaeology*, edited by J. Moore and E. Scott.

Flavell, John H. 1999. Cognitive Development: Children's Knowledge About the Mind. *Annual Review of Psychology* 50:21-45.

Gabbard, C. 2009. Studying action representation in children via motor imagery. *Brain Cogn* 71 (3):234-9.

Gamble, Clive, and Martin Porr. 2005. From empty spaces to lived lives - Exploring the individual in the Paleolithic. In *Hominid Individual in Context: Archaeological Investigations of Lower and Middle Paleolithic Landscapes, Locales and Artegacts*, edited by C. Gamble and M. Porr: Routledge.

Geribàs, Núria, Marina Mosquera, and Josep Maria Vergès. 2010. What novice knappers have to learn to become expert stone toolmakers. *Journal of Archaeological Science* 37 (11):2857-2870.

Gibson, J.J. 1979. *The ecological approach to visual perception*. Boston, Massachusetts: Houghton Mifflin.

Gill, S. V., Z. Yang, and Y. C. Hung. 2017. Effects of singular and dual task constraints on motor skill variability in childhood. *Gait Posture* 53:121-126.

Goldfield, Eugene C., Bruce A. Kay, and William H. Warren. 1993. Infant Bouncing: The Assembly and Tuning of Action Systems. *Child Development* 64:1128-1142.

Goldstein, Steven T. 2018. Knowledge Transmission Through the Lens of Lithic Production: a Case Study from the Pastoral Neolithic of Southern Kenya. *Journal of Archaeological Method and Theory* 26 (2):679-713.

Gonzalez-Forero, M., and A. Gardner. 2018. Inference of ecological and social drivers of human brain-size evolution. *Nature* 557 (7706):554-557.

Gowlett, J.A.J. 2005. Seeking the Paleolithic individula in East Africa and Europe during the Lower-Middle Pleistocene. In *Hominid Individual in Context: Archaeological Investigations of Lower and Middle Paleolithic Landscapes, Locales and Artegacts*, edited by C. Gamble and M. Porr: Routledge.

Repeated Author. 2006. The elements of design form in Acheulean bifaces: Modes, modalities, rules and language. In *Axe Age: Acheulean Tool-making from Quarry to Discard*, edited by N. Goren-Inbar and G. Sharon. London.

Guida, A., F. Gobet, H. Tardieu, and S. Nicolas. 2012. How chunks, long-term working memory and templates offer a cognitive explanation for neuroimaging data on expertise acquisition: a two-stage framework. *Brain Cogn* 79 (3):221-44.

Gurven, M. D., R. J. Davison, and T. S. Kraft. 2020. The optimal timing of teaching and learning across the life course. *Philos Trans R Soc Lond B Biol Sci* 375 (1803):20190500.

Häger-Ross, C., and B. Rösblad. 2002. Norms for grip strength in children aged 4-16 years. *Acta Pædiatr* 91:617-625.

Hasson, U., A. A. Ghazanfar, B. Galantucci, S. Garrod, and C. Keysers. 2012. Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends Cogn Sci* 16 (2):114-21.

Hecht, E. E., D. A. Gutman, N. Khreisheh, S. V. Taylor, J. Kilner, A. A. Faisal, B. A. Bradley, T. Chaminade, and D. Stout. 2015. Acquisition of Paleolithic toolmaking abilities involves structural remodeling to inferior frontoparietal regions. *Brain Struct Funct* 220 (4):2315-31.

Hecht, E. E., D. A. Gutman, T. M. Preuss, M. M. Sanchez, L. A. Parr, and J. K. Rilling. 2013. Process versus product in social learning: comparative diffusion tensor imaging of neural systems for action execution-observation matching in macaques, chimpanzees, and humans. *Cereb Cortex* 23 (5):1014-24.

Henrich, J., S. J. Heine, and A. Norenzayan. 2010. The weirdest people in the world? *Behav Brain Sci* 33 (2-3):61-83; discussion 83-135.

Hewlett, B. S., and C. J. Roulette. 2016. Teaching in hunter-gatherer infancy. *R Soc Open Sci* 3 (1):150403.

Hockett, Charles F. 1960. The Origin of Speech. *Scientific American* 203:88-97.

Högberg, Anders. 2008. Playing with Flint: Tracing a Child's Imitation of Adult Work in a Lithic Assemblage. *Journal of Archaeological Method and Theory* 15 (1):112-131.

Repeated Author. 2018. Approaches to Children's Knapping in Lithic Technology Studies. *Revista de Arqueologia* 31:58-74.

Holloway, R. L. 1966. Cranial Capacity, Neural Reorganization, and Hominid Evolution: A Search for More Suitable Parameters. *American Anthropologist* 68 (1):103-121.

Holloway, R. L., D. C. Broadfield, and M. S. Yuan. 2003. Morphology and histology of chimpanzee primary visual striate cortex indicate that brain reorganization predated brain expansion in early hominid evolution. *Anat Rec A Discov Mol Cell Evol Biol* 273 (1):594-602.

Hopkinson, Terry, and Mark J. White. 2005. The Acheulean and the handaxe: structure and agency in the Paleolithic. In *Hominid Individual in Context: Archaeological Investigations of Lower and Middle Paleolithic Landscapes, Locales and Artegacts*, edited by C. Gamble and M. Porr: Routledge.

Hrdy, S. B. 2009. *Mothers and Others: The Evolutionary Origins of Mutual Understanding*. Cambridge, MA: Belknap Press.

Humphrey, L. T. 2010. Weaning behaviour in human evolution. *Semin Cell Dev Biol* 21 (4):453-61.

Hunt, G. R., and R. D. Gray. 2003. Diversification and cumulative evolution in New Caledonian crow tool manufacture. *Proc Biol Sci* 270 (1517):867-74.

Isaac, Glynn. 1977. *Olorgesailie: Archaeological Studies of a Middle Pleistocene Lake Basin in Kenya*: University of Chicago Press.

Jerison, Harry J. 1979. Brain, Body and Encephalization in Early Primates. *Journal of Human Evolution* 8:615-635.

Johansen, L., and Dick. Stapert. 2012. Some workpieces by Middle Paleolithic apprentice flintknappers from the Netherlands. In *A mind set on flint: Studies in honour of Dick Stapert*, edited by M. J. L. T. Niekus, R. N. E. Barton, M. Street and T. Terberger.

Johansson, Niklas, Andrey Anikin, and Nikolay Aseyev. 2019. Color sound symbolism in natural languages. *Language and Cognition* 12 (1):56-83.

Kamp, Kathryn A. 2015. Children and their Childhoods: Retrospectives and Prospectives. *Childhood in the Past* 8 (2):161-169.

Karlin, Claudine, and Michèle Julien. 2019. An autumn at Pincevent (Seine-et-Marne, France): refitting for an ethnographic approach of a Magdalenian settlement. *Archaeological and Anthropological Sciences* 11 (9):4437-4465.

Kaschak, Michael P., and Arthur M. Glenberg. 2000. Constructing Meaning: The Role of Affordances and Grammatical Constructions in Sentence Comprehension. *Journal of Memory and Language* 43 (3):508-529.

Kawai, Masao. 1965. Newly-acquired Pre-cultural Behavior of the Natural Troop of Japanese Monkeys on Koshima Islet. *Primates* 6.

Kay, Paul, and Willett Kempton. 1984. What is the Sapir-Whorf Hypothesis? *American Anthropologist* 86:65-79.

Key, A., T. Proffitt, and I. de la Torre. 2020. Raw material optimization and stone tool engineering in the Early Stone Age of Olduvai Gorge (Tanzania). *J R Soc Interface* 17 (162):20190377.

Kline, Michelle Ann. 2016. TEACH: An Ethogram-based Method to Observe and Record Teaching Behavior. *Field Methods* 29 (3):205-220.

Knight, Charles L. F. 2017. Children, apprenticeship and pedagogy: Domestic crafting and obsidian core production at the Zaragoza-Oyameles source area in Puebla, Mexico. *Journal of Anthropological Archaeology* 47:152-166.

Konner, Melvin. 2010. *The Evolution of Childhood*. Cambridge, Massachusetts: Belknap Press.

Repeated Author. 2016. Hunter-Gatherer Childhood in the Context of Human Evolution. In *Childhood: Origins, Evolution, & Implications*, edited by C. L. Meehan and A. N. Crittenden. Albuquerque: University of New Mexico Press.

Kuhn, Deanna, and Maria Pease. 2006. Do Children and Adults Learn Differently? *Journal of Cognition and Development* 7 (3):279-293.

Lagattuta, K. H., H. J. Kramer, K. Kennedy, K. Hjortsvang, D. Goldfarb, and S. Tashjian. 2015. Beyond Sally's missing marble: further development in children's understanding of mind and emotion in middle childhood. *Adv Child Dev Behav* 48:185-217.

Lakoff, George, and Mark Johnson. 1980. *Metaphors We Live By*: University of Chicago Press.

Laland, K. N. 2017. The origins of language in teaching. *Psychon Bull Rev* 24 (1):225-231.

Laland, K. N., T. Uller, M. W. Feldman, K. Sterelny, G. B. Muller, A. Moczek, E. Jablonka, and J. Odling-Smee. 2015. The extended evolutionary synthesis: its structure, assumptions and predictions. *Proc Biol Sci* 282 (1813):20151019.

Leadbeater, E., N. E. Raine, and L. Chittka. 2006. Social learning: ants and the meaning of teaching. *Curr Biol* 16 (9):R323-5.

Leakey, Mary Douglas. 1971. *Olduvai Gorge: Volume 3, excavations in beds I and II, 1960-1963*. Vol. 3: Cambridge University Press.

Leigh, S. R. 2004. Brain growth, life history, and cognition in primate and human evolution. *Am J Primatol* 62 (3):139-64.

Leon-Carrion, J., J. Garcia-Orza, and F. J. Perez-Santamaria. 2004. Development of the inhibitory component of the executive functions in children and adolescents. *Int J Neurosci* 114 (10):1291-311.

Li, Li, Sam C. Lin, Shannon P. McPherron, Aylar Abdolahzadeh, Annie Chan, Tamara Dogandžić, Radu Iovita, George M. Leader, Matthew Magnani, Zeljko Rezek, and Harold L. Dibble. 2022. A Synthesis of the Dibble et al. Controlled Experiments into the Mechanics of Lithic Production. *Journal of Archaeological Method and Theory*.

Lillehammer, Grete. 1989. A Child is Born: The Child's World in an Archaeological Perspective. *Norwegian Archaeological Review* 22:89-105.

Repeated Author. 2015. 25 Years with the 'Child' and the Archaeology of Childhood. *Childhood in the Past* 8 (2):78-86.

Lombao, D., M. Guardiola, and M. Mosquera. 2017. Teaching to make stone tools: new experimental evidence supporting a technological hypothesis for the origins of language. *Sci Rep* 7 (1):14394.

Maloney, Tim Ryan. 2019. Towards Quantifying Teaching and Learning in Prehistory Using Stone Artifact Reduction Sequences. *Lithic Technology* 44 (1):36-51.

Mann, J., M. A. Stanton, E. M. Patterson, E. J. Bienenstock, and L. O. Singh. 2012. Social networks reveal cultural behaviour in tool-using dolphins. *Nat Commun* 3:980.

McCormick, Kelly, Jee Young Kim, Sara List, and Lynne C. Nygaard. 2015. Sound to Meaning Mappings in the Bouba-Kiki Effect. *CogSci*:1565-1570.

Migliano, A. B., and L. Vinicius. 2022. The origins of human cumulative culture: from the foraging niche to collective intelligence. *Philos Trans R Soc Lond B Biol Sci* 377 (1843):20200317.

Miller, I. F., S. E. Churchill, and C. L. Nunn. 2019. Speeding in the slow lane: Phylogenetic comparative analyses reveal that not all human life history traits are exceptional. *J Hum Evol* 130:36-44.

Moore, Mark W. 2011. The design space of stone flaking: implications for cognitive evolution. *World Archaeology* 43 (4):702-715.

Moore, Richard. 2013. Social learning and teaching in chimpanzees. *Biology & Philosophy* 28 (6):879-901.

Morgan, T. J., N. T. Uomini, L. E. Rendell, L. Chouinard-Thuly, S. E. Street, H. M. Lewis, C. P. Cross, C. Evans, R. Kearney, I. de la Torre, A. Whiten, and K. N. Laland. 2015. Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nat Commun* 6:6029.

Muller, A., C. Shipton, and C. Clarkson. 2022. Stone toolmaking difficulty and the evolution of hominin technological skills. *Sci Rep* 12 (1):5883.

Nahhas, R. W., A. C. Choh, M. Lee, W. M. Chumlea, D. L. Duren, R. M. Siervogel, R. J. Sherwood, B. Towne, and S. A. Czerwinski. 2010. Bayesian longitudinal plateau model of adult grip strength. *Am J Hum Biol* 22 (5):648-56.

Nakahashi, Wataru, Shiro Horiuchi, and Yasuo Ihara. 2018. Estimating hominid life history: the critical interbirth interval. *Population Ecology* 60 (1-2):127-142.

Navarrete, A., C. P. van Schaik, and K. Isler. 2011. Energetics and the evolution of human brain size. *Nature* 480 (7375):91-3.

Neill, D. 2014. Evolution of lifespan. *J Theor Biol* 358:232-45.

Neubauer, Fernanda. 2018. Playing with Projectile Points: Childhood Flintknapping Imitation at an 8,500-Year-Old Hunter-Gatherer Rockshelter Site in Southern Brazil. *Revista de Arqueologia* 31 (2):35-57.

Repeated Author. 2018. Playing with projectile points: Childhood flintknapping imitation at an 8,500 Year Old Hunter-Gatherer rockshelter site in southern Brazil. *Revista de Arqueologia* 31 (2):35-57.

Neubauer, S. 2014. Endocasts: possibilities and limitations for the interpretation of human brain evolution. *Brain Behav Evol* 84 (2):117-34.

Newcomer, M.H. 1971. Some Quantitative Experiments in Handaxe Manufacture. *World Archaeology* 3:85-104.

Newell, A., and P.S. Rosenbloom. 1981. Mechanisms of skill acquisition and the law of practice. In *Cognitive skills and their acquisition*, edited by J. R. Anderson. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Norman, Joel. 2002. Two visual systems and two theories of perception: An attempt to reconcile the constructivist and ecological approaches. *Behavioral and Brain Sciences* 25:73-144.

Nowell, April, and Jennifer C. French. 2020. Adolescence and innovation in the European Upper Palaeolithic. *Evolutionary Human Sciences* 2.

Odell, George. 1981. The Mechanics of Use-Breakage of Stone Tools: Some Testable Hypotheses. *Journal of Field Archaeology* 8:197-209.

Osiurak, F., Y. Rossetti, and A. Badets. 2017. What is an affordance? 40 years later. *Neurosci Biobehav Rev* 77:403-417.

Pargeter, J., N. Khreisheh, and D. Stout. 2019. Understanding stone tool-making skill acquisition: Experimental methods and evolutionary implications. *J Hum Evol* 133:146-166.

Piaget, Jean. 1952. *The Origins of Intelligence in Children*. Translated by M. Cook. New York: International Universities Press, Inc.

Piaget, Jean, and Barbel Inhelder. 1956. *The child's conception of space*. London: Routledge & Kegan Paul.

Pope, Matt, and Mark Roberts. 2005. Observations on the relationship between Palaeolithic individuals individual and artefact scatters at the Middle Pleistocene site of Boxgrove, UK. In *Hominid Individual in Context: Archaeological Investigations of Lower and Middle Paleolithic Landscapes, Locales and Artegacts*, edited by C. Gamble and M. Porr: Routledge.

Putt, S. S. J., S. Wijeakumar, and J. P. Spencer. 2019. Prefrontal cortex activation supports the emergence of early stone age toolmaking skill. *Neuroimage* 199:57-69.

Putt, Shelby S., Sobanawartiny Wijeakumar, Robert G. Franciscus, and John P. Spencer. 2017. The functional brain networks that underlie Early Stone Age tool manufacture. *Nature Human Behaviour* 1 (6).

Rabaglia, C. D., S. J. Maglio, M. Krehm, J. H. Seok, and Y. Trope. 2016. The sound of distance. *Cognition* 152:141-149.

Randerath, J., G. Goldenberg, W. Spijkers, Y. Li, and J. Hermsdorfer. 2011. From pantomime to actual use: how affordances can facilitate actual tool-use. *Neuropsychologia* 49 (9):2410-6.

Richerson, Peter J., and Robert Boyd. 2005. *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press.

Riede, F., N. N. Johannsen, A. Hogberg, A. Nowell, and M. Lombard. 2018. The role of play objects and object play in human cognitive evolution and innovation. *Evol Anthropol* 27 (1):46-59.

Roux, V., B. Bril, and G. Dietrich. 1995. Skills and learning difficulties involved in stone knapping: The case of stone-bead knapping in Khambhat, India. *World Archaeology* 27:63-87.

Sandgathe, Dennis M. 2017. Identifying and Describing Pattern and Process in the Evolution of Hominin Use of Fire. *Current Anthropology* 58 (S16):S360-S370.

Satta, E., S. Ferrari-Toniolo, F. Visco-Comandini, R. Caminiti, and A. Battaglia-Mayer. 2017. Development of motor coordination during joint action in mid-childhood. *Neuropsychologia* 105:111-122.

Saxena, S., E. Cinar, A. Majnemer, and I. Gagnon. 2017. Does dual tasking ability change with age across childhood and adolescence? A systematic scoping review. *Int J Dev Neurosci* 58:35-49.

Schaefer, S., D. Jagenow, J. Verrel, and U. Lindenberger. 2015. The influence of cognitive load and walking speed on gait regularity in children and young adults. *Gait Posture* 41 (1):258-62.

Shea, J. J. 2006. Child's Play: Reflections on the Invisibility of Children in the Paleolithic Record. *Evolutionary Anthropology* 15:212-216.

Repeated Author. 2017. Occasional, obligatory, and habitual stone tool use in hominin evolution. *Evol Anthropol* 26 (5):200-217.

Shinohara, Kazuko, and Shigeto Kawahara. 2010. A Cross-linguistic Study of Sound Symbolism: The Images of Size. *Annual Meeting of the Berkeley Linguistics Society* 36 (1).

Shultz, S., E. Nelson, and R. I. Dunbar. 2012. Hominin cognitive evolution: identifying patterns and processes in the fossil and archaeological record. *Philos Trans R Soc Lond B Biol Sci* 367 (1599):2130-40.

Sidhu, D. M., C. Westbury, G. Hollis, and P. M. Pexman. 2021. Sound symbolism shapes the English language: The maluma/takete effect in English nouns. *Psychon Bull Rev* 28 (4):1390-1398.

Simpson, George Gaylord. 1953. The Baldwin Effect. *Evolution* 7 (2):110-117.

Smaers, J. B., and C. Soligo. 2013. Brain reorganization, not relative brain size, primarily characterizes anthropoid brain evolution. *Proc Biol Sci* 280 (1759):20130269.

Smith, Linda, and Michael Gasser. 2005. The Development of Embodied Cognition: Six Lessons from Babies. *Artificial Life* 11:13-29.

Smits-Engelsman, B. C., D. Sugden, and J. Duysens. 2006. Developmental trends in speed accuracy trade-off in 6-10-year-old children performing rapid reciprocal and discrete aiming movements. *Hum Mov Sci* 25 (1):37-49.

Sporns, O., and R. F. Betzel. 2016. Modular Brain Networks. *Annu Rev Psychol* 67:613-40.

Spruijt, S., J. van der Kamp, and B. Steenbergen. 2015. The ability of 6- to 8-year-old children to use motor imagery in a goal-directed pointing task. *J Exp Child Psychol* 139:221-33.

Sterelny, K. 2011. From hominins to humans: how sapiens became behaviourally modern. *Philos Trans R Soc Lond B Biol Sci* 366 (1566):809-22.

Sterelny, Kim. 2017. Artifacts, Symbols, Thoughts. *Biological Theory* 12 (4):236-247.

Sternke, F., and M. Sørensen. 2005. The identification of children's flintknapping products in Mesolithic Scandinavia. In *Mesolithic Horizons*, edited by S. McCartan, R. Schulting, G. Warren and P. Woodman. Oxford: Oxbow Books.

Stout, D. 2002. Skill and Cognition in Stone Tool Production: An Ethnographic Case Study from Irian Jaya. *Current Anthropology* 43:693-722.

Stout, D., and T. Chaminade. 2007. The evolutionary neuroscience of tool making. *Neuropsychologia* 45 (5):1091-100.

Stout, D., E. Hecht, N. Khreisheh, B. Bradley, and T. Chaminade. 2015. Cognitive demands of lower paleolithic toolmaking. *PLoS One* 10 (4):e0121804.

Stout, D., R. Passingham, C. Frith, J. Apel, and T. Chaminade. 2011. Technology, expertise and social cognition in human evolution. *Eur J Neurosci* 33 (7):1328-38.

Stout, D., N. Toth, and K. Schick. 2006. Comparing the neural foundations of Oldowan and Acheulean toolmaking: A pilot study using positron emission tomography (PET). In *The Oldowan: Case Studies in the Earliest Stone Age*, edited by N. Toth and K. Schick. Gosport: Stone Age Institute Press.

Stout, D., N. Toth, K. Schick, and T. Chaminade. 2008. Neural correlates of Early Stone Age toolmaking: technology, language and cognition in human evolution. *Philos Trans R Soc Lond B Biol Sci* 363 (1499):1939-49.

Stout, Dietrich. 2011. Stone toolmaking and the evolution of human culture and cognition. *Philos Trans R Soc Lond B Biol Sci* 366 (1567):1050-9.

Stout, Dietrich, and Nada Khreisheh. 2015. Skill Learning and Human Brain Evolution: An Experimental Approach. *Cambridge Archaeological Journal* 25 (4):867-875.

Stout, Dietrich, Nicholas Toth, Kathy Schick, Julie Stout, and Gary Hutchins. 2000. Stone Tool-Making and Brain Activation: Position Emission Tomography (PET) Studies. *Journal of Archaeological Science* 27 (12):1215-1223.

Street, S. E., A. F. Navarrete, S. M. Reader, and K. N. Laland. 2017. Coevolution of cultural intelligence, extended life history, sociality, and brain size in primates. *Proc Natl Acad Sci U S A*.

Sullivan, Katherine J., Shailesh S. Kantak, and Patricia A. Burtner. 2008. Motor Learning in Children: Feedback Effects on Skill Acquisition. *Physical Therapy* 88:720-732.

Tennie, C., J. Call, and M. Tomasello. 2009. Ratcheting up the ratchet: on the evolution of cumulative culture. *Philos Trans R Soc Lond B Biol Sci* 364 (1528):2405-15.

Thompson, J. L., and A. J. Nelson. 2011. Middle childhood and modern human origins. *Hum Nat* 22 (3):249-80.

Toth, N. 1985. The Oldowan Reassessed: A Close Look at Early Stone Artifacts. *Journal of Archaeological Science* 12:101-120.

Toth, Nicholas, and Kathy Schick. 2018. An overview of the cognitive implications of the Oldowan Industrial Complex. *Azania: Archaeological Research in Africa* 53 (1):3-39.

Tringham, Ruth, Glenn Cooper, George Odell, Barbara Voytek, and Anne Whitman. 1974. Experimentation in the Formation of Edge Damage: A New Approach to Lithic Analysis. *Journal of Field Archaeology* 1:171-196.

Tsur, Reuven. 2006. Size–sound symbolism revisited. *Journal of Pragmatics* 38 (6):905-924.

Vainio, L., M. Vainio, J. Lipsanen, and R. Ellis. 2019. The Sound of Grasp Affordances: Influence of Grasp-Related Size of Categorized Objects on Vocalization. *Cogn Sci* 43 (10):e12793.

Vančata, Václav. 1996. Major Patterns of Early Hominid Evolution: Body Size, Proportions, Encephalization and Sexual Dimorphism. *Anthropologie* 34:11-25.

Vygotsky, L.S. 1978. *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Vygotsky, Lev. 1978. Interaction Between Learning and Development. In *Mind and Society*. Cambridge, MA: Harvard University Press.

Walker, R., O. Burger, J. Wagner, and C. R. Von Rueden. 2006. Evolution of brain size and juvenile periods in primates. *J Hum Evol* 51 (5):480-9.

Wang, J. J., M. Ali, S. Frisson, and I. A. Apperly. 2016. Language complexity modulates 8- and 10-year-olds' success at using their theory of mind abilities in a communication task. *J Exp Child Psychol* 149:62-71.

Washburn, S.L., and C.S. Lancaster. 1968. The Evolution of Hunting. In *Man the hunter*, edited by R. B. Lee and I. DeVore.

Weisner, T.S. 1984. Ecocultural Niches of Middle Childhood: A Cross-Cultural Perspective. In *Development During Middle Childhood: The Years from Six to Twelve*, edited by W. A. Collins. Washington, D.C.: National Academy Press.

Repeated Author. 1996. The 5 to 7 transition as an ecocultural project. In *The five to seven year shift: The age of reason and responsibility*, edited by A. J. Sameroff and M. M. Haith: University of Chicago Press.

Wenner, Adrian M. 1962. Sound Production During the Waggle Dance of the Honey Bee. *Animal Behaviour* 10:79-95.

Whittaker, John C. 2004. *American Flintknappers*. Austin: University of Texas Press.

Whorf, Benjamin Lee. 1988. The Relation of Habitual Thought and Behavior to Language. In *High Points in Anthropology*, edited by P. Bohannan and M. Glazer: McGraw-Hill, Inc.

Wilson, Margaret. 2002. Six views of embodied cognition. *Psychonomic Bulletin & Review* 9:625-636.

Wynn, T. 2002. Archaeology and cognitive evolution. *Behavioral and Brain Sciences* 25:389-438.

Wynn, T., and F. L. Coolidge. 2016. Archeological insights into hominin cognitive evolution. *Evol Anthropol* 25 (4):200-13.

Wynn, T., R. A. Hernandez-Aguilar, L. F. Marchant, and W. C. McGrew. 2011. "An ape's view of the Oldowan" revisited. *Evol Anthropol* 20 (5):181-97.

Zipp, G.P., and A.M. Gentile. 2010. Practice Schedule and the Learning of Motor Skills in Children and Adults: Teaching Implications. *Journal of College Teaching and Learning* 7:35-42.

CHAPTER 2: METHODS AND RESULTS

METHODS

This study examines Oldowan-style, least effort stone toolmaking skill acquisition across two developmental groups: children (ages 8 to 13 years) and adults (ages 18 years and older). All participants, regardless of experimental condition, engaged in five consecutive days of stone knapping training, with the thesis writer (MBK) acting as the instructor. Training consisted of one 30-minute practice session per day (2.5 hours, total), which included both direct-active instruction and time for the participants to practice independently and ask questions. For adults, training took place at the Emory University Paleolithic Technology Laboratory. For children, the study was held in conjunction with an archaeology-themed science summer education program. Children who enrolled in the archaeology program ("Archaeology Adventure Week") were not required to also enroll in the study, as the study and the program were separate entities. All activities, including those that were specific to the study (e.g. psychometrics), were completed by all children who signed up for the summer program to ensure that anyone enrolling in the summer program would not feel pressure to also engage in the study.

Data were collected between June and December 2021. For the child participants, data were collected during two one-week sessions of the Tellus Museum's Archaeology Adventure Week summer program. The adults were recruited and sessions were held during the Emory University Fall 2021 semester.

Participant Recruitment, Informed Consent, and Compensation

A total of 50 participants were recruited to participate in the study: 20 children (Male: n = 11; Female: n = 9; Median Age: 9 years) and 30 adults (Male: n = 8; Female: n = 21; Nonbinary:

n = 1; Median Age: 19 years). Of the children, one participant was excluded from the sample due to a motor disability that prevented them from being able to knap without the direct assistance of an adult, reducing the number of child participants to 19 (Male: n = 11; Female: n = 8; Median Age: 9 years). Both groups of participants were recruited from approved flyers, email advertisements, and through word-of-mouth. Children were additionally recruited through the Tellus Science Museum's advertising department, via the museum's website and membership newsletter, as a part of the co-occurring archaeology-themed summer program.

At the beginning of their first training session, all participants were guided through the informed consent process by MBK. Adult participants completed these forms independently, whereas the children's legal guardians provided consent in their place. Children participants went through an additional assent process – those between the ages of 7 and 10 years indicated their assent verbally, and those between the ages of 11 and 13 years provided their signature. All participants were also asked to review and indicate their consent to having video data with their likeness published on a data sharing website, Databrary.

Due to the study's association with the archaeology-themed summer program, neither child participants nor their legal guardians were compensated financially for the children's participation in the study. To incentivize retention, adult participants were awarded $10 per practice session, with a $20 bonus for study completion, for a total of $70. Funds were dispersed to participants using the Emory University ClinCard system.

Core Material, Preparation, and Selection

Prior to the experiment, unmodified basalt nodules were spalled into smaller cores using an 8 lb. sledgehammer and sorted into one of the following size categories: small (400 – 699g), medium (700 – 999g) and large (1000 – 1300g). The material was sourced from neolithics.com and was chosen due to its similarity to the stone materials used by ancient Oldowan knappers in the Afar region of Gona, Ethiopia (Stout et al. 2005). Size categories were determined from the instructor's best estimate of children's and adults' ability to handle cores of different sizes, the reasoning being that the small cores may be a comfortable fit for children's hands but too small to easily manipulate for the adults, that the large cores may be a comfortable fit for adults' hands but too large to easily manipulate for the children, and that the medium cores may be comfortable for both groups. The estimated degree of comfort with various core sizes was based on the instructor's own experience with Oldowan-style stone toolmaking.

Because the use of a sledgehammer during spalling yields cores of unpredictable sizes and shapes, cores were selected from the resulting fractured pieces not only along the criterion of weight, but also of approximate shape. Cores that were determined to be too flat and thin (i.e. like a tablet) or too round (i.e. like a ball) were excluded from the sample, the former because such a shape may provide an unfair advantage to the participant who received it (e.g. the core potentially would contain an unusually high number of viable platform angles, and the greater surface area could result in larger, more expert-looking flakes without reflecting accurately the skill of the knapper) and the latter because such a shape would be disadvantageous to the participant who received it (e.g. the core would have a paucity of viable platform angles, and it would be difficult for knappers, especially novices, to create opening flakes). Core shapes that were selected for the experiment fell between these two extremes.

Upon selection, linear measurements (maximum length, width, and thickness) and starting weight were recorded for each core, and all cores were numbered and spray-painted a silver color to impose an artificial cortex, given that the stone's natural cortex was removed during the spalling process. Silver was selected as the paint color so that debitage produced during the experiment could be more easily photographed, while still somewhat resembling the natural grey of the basalt.

All participants were randomly assigned a core from each size category for each day of the experiment, such that every day, they had one small, one medium, and one large core, for a total of three cores in their session set. A complete new set of cores was given to each participant at the beginning of each practice session, regardless of whether they knapped all three of their cores on the previous day. In the interest of raw material conservation, cores unused by one participant were permitted to be reassigned to another participant on a subsequent day, but a single participant did not encounter the same unused core more than once.

During the sessions, participants were instructed to knap the cores in the order of their choosing and were advised that once they began working on a core, they would be required to knap the core to "completion" before moving on to the next one. "Completion" criteria, defined for this experiment, were met when one or both of the two following conditions were satisfied: 1) the participant reduced the core to a size they felt was no longer safe to knap (i.e. too small) and/or 2) the participant eliminated all viable angles on the core and was no longer finding success with flake removal. If a core broke into two roughly equal pieces during knapping (i.e. a "core split"), participants were told to choose one half to continue knapping as the core and to place the other half with the other detached pieces. Participants were not allowed to return to the unused half of the core split after exhausting the chosen half. Once participants had finished with a core, they were instructed to place all debitage in a provided bag and the core in a separate bag, for later analysis.

Hammerstone Material and Selection

Hammerstones used during the experiment were “Mexican beach pebbles,” a type of basalt landscaping material sourced from a chain home improvement store, Home Depot. From the purchased materials, stones that were roughly ovoid in shape and between 9 and 15 cm were included in the hammerstone sample. During the training sessions, participants were instructed to select a hammerstone about the size of their palm that felt comfortable in their dominant hand. If participants were uncertain about which hammerstone to choose, the instructor would make a recommendation. All participants in a session had access to the same selection of hammerstones. If a hammerstone broke during knapping or developed wear patterns that made it unusable, it was removed from the experiment. Participants had the option to switch hammerstones at any time during the training sessions (e.g. if their hammerstone broke, they were having difficulty removing flakes, or at the instructor’s suggestion).

Training Session Setup, Knapping Instructions, and Procedures

Upon recruitment, participants were advised to come to their sessions wearing closed-toe shoes and long pants, to protect their feet and legs from sharp debitage. At the beginning of each session, participants were given safety glasses and cut-resistant gloves. Adults were given the option of also using a leather lap mat, should they choose to stabilize the core on their leg. For children, the use of the lap mat was required.

At each training session, chairs equal to the number of participants plus the instructor were arranged in a circle, with the instructor’s chair positioned to be visible to all participants. Weather permitting, the sessions took place outside. On rainy days, adult participants knapped inside the Paleolithic Technology Laboratory – bad weather was no obstacle for the child participants, as their sessions took place on a covered patio. All training sessions were video recorded with a Sony – Handycam CX405. For groups of larger than three participants, two cameras were used and positioned to capture footage of all participants in the session.

On Day 1, prior to knapping, participants were given a demonstration of least effort, Oldowan-style stone toolmaking, along with a description of the goal of the study (i.e. that the participants would learn to make stone flakes), and instructions on how to knap. Oldowan stone toolmaking has been described as a ‘least effort’ approach to knapping because the process likely required ancient knappers to utilize the least amount of time and energy to produce a sharp edge (Isaac and Harris 1997 in Braun 2011). Knapping was primarily performed unifacially, and the desired product was most likely a stone flake that required little to no refinement once made. This is in contrast with later stone tool technologies, such as the Acheulean, that were knapped bifacially, exploited flake scars for successive percussions, and whose product was planned and shaped (cite).

The instructor emphasized the importance of the following: (1) identifying and exploiting acute platform angles, (2) proper posture during knapping (i.e. sitting upright with feet about shoulder-width apart), and (3) correct hammerstone grip and usage (i.e. holding the hammerstone “like a ball” and striking the core with follow-through). Participants were also told that there were two ways to properly hold the core (i.e. freehand or supported on their leg), and that depending upon the core’s size, they may want to favor one posture over the other. After producing a whole flake during the demonstration, the instructor showed the flake to the participants and identified the relevant features that classified it as a “good” or “useful” flake (e.g. size, cutting edge, thickness).

Following the demonstration, participants were instructed to select their first core. If participants struggled with this decision, the instructor recommended that they look for a core with characteristics like those described during the demonstration (e.g. a core with identifiable acute angles). After selecting a core, participants were told to write the core number down on a provided form and to set aside the bag with the corresponding number, for later use. Participants were then instructed to try to make flakes for the next 30 minutes or until they had exhausted all three of their provided cores.

During the training sessions, participants could ask questions of the instructor at any time. Instruction was unlimited, in the form of direct-active teaching (Kline 2016) and could address any subject related to knapping. If a participant appears to be having difficulty removing flakes, the instructor would offer suggestions and remind the participant of the principles of knapping introduced during the demonstration. Such suggestions included repositioning the participant's posture, remembering to knap with follow-through, and looking for acute angles on the core. On occasion (e.g. if the participant asked for more direct guidance), the instructor would take either the core or the hammerstone from the participant and pantomime the recommended behavior or adjustments. The instructor also made comments about safety protocols, such as adjusting a participant's grip on the core so they avoided hitting their fingers, and offered praise when participants displayed good knapping technique.

Days 2 – 5 followed the same procedure, without the extended demonstration period held on Day 1. Participants selected a hammerstone, chose the order in which to knap their three preselected cores, and attempted to remove flakes, with assistance and guidance from the instructor.

Participant group sizes varied across sessions due to scheduling availability and ranged from 2 to 6 in children (mean = 5.15, median = 6, mode = 6) and 1 to 4 in adults (mean = 1.43, median = 1, mode = 1). Additionally, given that some children who enrolled in the Tellus Museum's summer program chose not to participate in the study, children participants who were placed in groups with non-participants were exposed to larger group sizes. Participating children and non-participating children could not be divided into separate groups due to the organization of the summer program and the staffing needs of the museum. Because of this, group sizes for child participants are more accurately reported with the number of non-participating children included (range: 4 to 7, mean = 6.25, median = 6.5, mode = 7).

If a session consisted of two or more participants, all participants were allowed to interact with each other as naturalistically as possible. Although most interactions between participants were social (i.e. unrelated to knapping), participants could ask each other questions about the knapping task or make remarks about the study, in general (e.g. about the psychometric tasks or study design).

Lithic Analysis

I recruited and trained a team of five undergraduate research assistants to aid with the lithic analysis. All research assistants were trained in lithic analysis on test lithic assemblages that I produced for the purposes of training. Although I did not conduct formal inter-rater reliability testing for the lithic data collection, research assistants were not permitted to advance to the participant data set until they identified all debitage types correctly and their test assemblage measurements fell within 5mm of my own. All assistants collected measurements on lithics produced by participants from both conditions, and together, we analyzed all lithic debitage according to the following protocol.

First, all detached pieces produced from a single core were weighed, and any under 5g were removed from the sample. The remaining pieces were then numbered and sorted into one of four categories: (1) whole flake (i.e. point of percussion and platform were present and identifiable and cutting edge was complete), (2) split (i.e. a whole flake broken vertically through the point of percussion in two or more pieces), (3) snap (i.e. a whole flake broken horizontally across the flake's medial surface), and (4) detached piece (i.e. debitage that had no identifiable point of percussion or platform). Maximum linear dimensions (length, width, and thickness) were recorded for all detached pieces.

Whole flakes were further subjected to technological linear dimensions, measured with respect to the point of percussion (e.g. technological length was measured from the point of percussion to the distal cutting edge, at 90^o from the platform). The exterior platform angle for whole flakes was calculated using a measurement from a modified set of calipers, and platforms were categorized into one of seven types: (1) plain flat (i.e. platform is a smooth, flat surface that spans roughly the entire width of the flake), (2) dihedral (i.e. two facets joined by an angle), (3) multifaceted (i.e. more than two facets joined by two or more angles), (4) focalized (i.e. platform is central to the flake, with edges extending down from either side), (5) punctiform (i.e. a single point surrounding the point of percussion), (6) linear (i.e. no measurable platform width), or (7) shattered (i.e. the platform is crushed or incomplete). Platforms were also inspected for the presence of cortex and measured for length (i.e. the maximum dimension on the platform) and width (i.e. taken at 90^o of the length).

The reduced cores were measured for their ending maximum linear dimensions and weight. These values were used to calculate cores' reduction intensity (*Table 2.1*). Using the cores' starting dimensions, geometric means were calculated for beginning maximum length, width, and thickness to create a scale-free measure. These values were then applied in a factor analysis to determine whether the cores' initial shapes could be sorted into discernable categories (Stout et al, 2019). Two factors were identified, explaining 78.1% of core shape variance (Factor 1: 44.7%, Factor 2: 33.3%). Factor 1 was characterized by large positive loadings on core length and width. Factor 2 largely captured variation in thickness. Because of this, cores with high Factor 1 scores were characterized as having a "flatter" shape, and cores with a high Factor 2 score were characterized as being "rounder" (*Table 2.1*)

Table 2.1: PCA Loadings of Core Shape

	Factor 1	**Factor 2**
Variance %	44.7	33.4
Interpretation	Flatness	Roundness
Scale-Free Beginning Length	0.815	-0.107
Scale-Free Beginning Width	0.819	0.014
Scale-Free Beginning Thickness	0.076	0.995

Using a modified version of the calculation described in Morgan et al. (2015) supplemental materials, a utility score was calculated for all detached pieces. This score, described by the authors as flake "quality," took into account the amount of cutting edge present on the detached pieces, relative to its mass, and the size and shape of the detached piece, such that pieces with a high cutting edge length were rewarded and pieces were penalized for being excessively small (*Table 2.2*). In addition to this score, further measures of utility were taken with respect to a participant's overall knapping performance (e.g. number of detached pieces and whole flakes produced). (See *Table 2.3* below for more details.)

Table 2.2: Defining and Calculating "Reduction Intensity" and "Utility"

Variable Terminology	**Definition**	**Calculation**
Reduction Intensity	The total weight (g) removed from the starting core, represented as a percentage	$\frac{Ending\ Weight}{Starting\ Weight} \times 100$
Utility	A measure of a detached piece's quality, taking into account its length of cutting edge, size, and shape, with penalties for small size and benefits for high cutting edge length	$\left(\frac{Flake\ Cutting\ Edge\ Mass}{Flake\ Mass^{1/3}}\right) \times \left(1 - EXP\left(-0.31 \times (0.1 \times Flake\ Diameter - 0.81)\right)\right)$

Psychometric and Motor Testing

To assess individual differences between participants, all subjects completed the following series of tasks designed to evaluate their psychometric, physical, and motivation-related abilities: (1) Because stone toolmaking is a vigorous physical activity, individuals with greater upper body muscular strength may have an advantage in a stone toolmaking task over those with less upper body strength. Grip strength, a common approximation of upper body muscular strength (Wind et al. 2010), was measured in kilograms using an electric hand dynamometer (Camry EH101). (2) Additionally, skilled stone toolmaking requires precision, as knappers must be able to hit a stone nodule at the desired platform with enough speed and force to detach a flake. Thus, during stone toolmaking, stone toolmakers must have control over their speed/accuracy tradeoff, a phenomenon known as "Fitts' Law" (Fitts 1954). Participants' motor accuracy was measured using a computerized version of Fitts' Law test (Source: http://depts.washington.edu/acelab/proj/fittsstudy/index.html), which requires participants to use a mouse to click as quickly as possible between two target ribbons of varying thickness and spacing. (3) Working memory, selective attention, and cognitive flexibility, each facets of cognitive executive function, have garnered interest as potentially important mental capacities for successful stone toolmaking (Stout et al. 2015). Participants' ability to rotate objects mentally, a skill that would be useful in core manipulation and reduction strategy planning, was measured using the classic Mental Rotation Task (MRT) (Vandenberg and Kuse 1978). (4) Selective attention and cognitive flexibility was measured using the Stroop Color-Word Test (Stroop 1935), a well-known paradigm that asks participants to inhibit cognitive interference during the simultaneous processing of two conflicting stimuli (e.g. correctly reading aloud the word "red," which is written in blue ink). Age-appropriate versions of this test were administered in pencil and paper format to participants in both groups. (5) Finally, participants completed the Big Five Inventory (BFI), a cross-culturally validated (Benet-Martínez and John 1998) personality inventory well-known for its persistence across an individual's lifetime (Soto et al. 2011), appropriate to their developmental level (i.e. the original version for adults and version 46A for children). Of the five factors measured by the BFI (Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism), positive scores in conscientiousness and openness have been demonstrated to be predictive of high academic achievement in children

(Caprara et al. 2011) and young adults (Busato et al. 1999). Therefore, these personality characteristics may indicate more broadly an individual's ability to engage in long-term learning of a task that is cognitively and, in the case of stone toolmaking, physically demanding.

Completion of these tasks was divided across the first four days of a participant's five-day recruitment period. Because the archaeology-themed summer program was organized in rotating stations, child participants may have completed the psychometrics either before or after their knapping training session, depending on their rotation schedule that day. All adult participants completed the day's psychometrics prior to the knapping task.

RESULTS

Lithics

Knapping Skill Metrics: Quantity, Quality, and Flaking Inefficiency

Calculating and Defining Variables

From the flake and core measurements recorded, several metrics were identified as being particularly relevant indicators of knapper skill: per core, (1) core reduction intensity, (2) total number of detached pieces, (3) average utility of detached pieces, (4) total utility of detached pieces, (5) total number of whole flakes, (6) average utility of whole flakes, (7) total utility of whole flakes, (8) average weight of whole flakes, and (9) total weight of whole flakes. These criteria were selected from the broader range of measurements taken because they are representative of both the goals of the experimental task (i.e. to produce "good quality" whole flakes and to completely reduce the starting core) and of what are assumed to be the goals of ancient Oldowan knappers (i.e. to produce whole flakes with a viable cutting edge and to make conservative and efficient use of raw materials) (Delagnes and Roche 2005; Semaw et al. 1997). For more details on variable definitions and inclusion criteria, please refer to *Table 2.3*.

Table 2.3: Knapping Skill Metric Variable Definitions and Inclusion Rationale

Variable Name	**Definition**	**Inclusion Rationale**
Reduction Intensity	The total weight (g) removed from the starting core, represented as a percentage	Novice knappers are especially prone to crushing viable platform angles, leaving the core "rounded-off." Once this rounding-off occurs, further flake extractions from the core become increasingly difficult, and cores are more likely to be abandoned prematurely (i.e. before they are completely reduced). Knappers of greater skill are better able to maintain viable platform angles and are thus more likely to reduce their cores to smaller sizes, as reflected in a higher percentage of reduction intensity.
Number of Detached Pieces - Total	The total number of detached pieces of all types (i.e. whole flakes, split flakes, snapped flakes, and shatter) produced by a single participant from all cores and practice sessions	This value is meant to account for instances of "core splits" (i.e. the core breaks roughly in half during knapping) and the removal of large shatter, both of which may artificially inflate the measure of a knapper's skill as estimated by Reduction Intensity. Although large whole flake size is typically an indicator of higher degrees of knapper skill, this fact must be counterbalanced by acknowledging that novices are also prone to producing large detached pieces of low utility. Because of this, it is hypothesized that knappers who extract more detached pieces of higher quality from a core are more likely to demonstrate better control over knapping fracture mechanics than those who remove a smaller number of larger pieces.
Utility of Detached Pieces - Average	The mean utility score of detached pieces of all types produced by a single participant from all cores and practice sessions	Although it is widely assumed that ancient hominin Oldowan knappers would have preferentially used and produced whole flakes, it is possible that, per the conservation of raw material, they also would have made use of detached pieces of a variety of types, so long as they still met the criteria of a useful tool (e.g. larger size and long, viable cutting edge). With this in mind, a utility score was calculated for all detached pieces, regardless of classified type. By averaging these scores for a single participant, we can examine how reliably a participant can produce a useful stone tool, irrespective of how many detached pieces they produced, in total.
Utility of Detached Pieces - Total	The sum of the utility scores of detached pieces of all types produced by a single participant from all cores and practice sessions	As with *Utility of Detached Pieces - Average* , but taking into consideration the total number of detached pieces a single participant produced.
Number of Whole Flakes - Total	The total number of all whole flakes produced by a single participant from all cores and practice sessions	The production of whole flakes, as opposed to split flakes, snapped flakes, or shatter, is indicative of a knapper's ability to: 1) identify and exploit viable platform angles, 2) correctly employ the hammerstone (e.g. in terms of angle of approach, velocity, and follow-through), and 3) properly support the core with the nondominant hand such that force is evenly distributed and the flake does not break into two or more pieces. The greater the number of whole flakes a participant produces, the more likely it is that they have become gained skill in these knapping principles.
Utility of Whole Flakes - Average	The mean utility score of all whole flakes produced by a single participant from all cores and practice sessions	The production of whole flakes of large size and long cutting edge length is assumed to be the primary goal of Oldowan knapping. Averaging utility score of a single participant's whole flakes allows us to examine how reliably participants were able to achieve this goal, irrespective of how many whole flakes they produced, in total.
Utility of Whole Flakes - Total	The sum of the utility scores of all whole flakes produced by a single participant from all cores and practice sessions	As with *Utility of Whole Flakes - Average* , but taking into consideration the total number of whole flakes a single participant produced.
Weight of Whole Flakes - Average	The mean weight (g) of all whole flakes produced by a single participant from all cores and practice sessions	Larger whole flakes are typically considered to be of greater utility than smaller flakes, as they are more likely to have a greater length of cutting edge. They are also more difficult to produce than smaller flakes, given that they require a knapper to strike a viable platform at the correct point on the core (i.e. near the platform edge, but not so close that the platform crumbles and becomes unusable) with a discerning amount of force (i.e. enough for the fracture to carry through to the other side of the core and create a sharp cutting edge but not so much that it produces an "overshot"). In combination, these factors are indicative of an experienced knapper who has gained at least some mastery over fracture mechanics. Averaging the weight of a single participant's whole flakes allows for an estimate of how reliably they were able to produce whole flakes of larger size, irrespective of how many whole flakes they produced, in total.
Weight of Whole Flakes - Total	The sum weight (g) of all whole flakes produced by a single participant from all cores and practice sessions	As with *Weight of Whole Flakes - Average* , but taking into consideration the total number of whole flakes a single participant produced.

These nine variables yielded three factors when entered into a principle components analysis (PCA), which used a correlation matrix with no rotation: Factor 1, with an eigenvalue of 50.0, which had high positive loadings on core reduction intensity, the total number and utility of detached pieces, and the total number, utility, and weight of whole flakes; Factor 2, with an eigenvalue of 18.2, which had high positive loadings on the average utility of detached pieces and the average weight and utility of whole flakes; and Factor 3, with an eigenvalue of 11.7, which negatively loaded on reduction intensity and had modest or marginal loadings on all other variables. Together, these factors account for 79.9% of variable variance. (See *Table 2.4*) Although Factors 2 and 3 demonstrated much lower eigenvalues than Factor 1, and were therefore more marginal measures of knapper skill, all three of the factors produced by the PCA were included in the analysis because each seemed to account for different aspects of participant lithic outcomes and knapper decision-making.

Table 2.4: PCA Loadings of Knapping Skill Metrics

	Factor 1	**Factor 2**	**Factor 3**
Variance %	50.0	18.2	11.7
Interpretation	Quantity	Quality	Flaking Inefficiency
Reduction Intensity	0.655	0.149	-0.608
Number of Detached Pieces - Total	0.859	-0.241	-0.312
Utility of Detached Pieces - Average	0.291	0.650	0.288
Utility of Detached Pieces - Total	0.895	-0.133	-0.218
Number of Whole Flakes - Total	0.809	-0.414	0.342
Utility of Whole Flakes - Average	0.520	0.601	0.357
Utility of Whole Flakes - Total	0.851	-0.295	0.397
Weight of Whole Flakes - Average	0.337	0.701	-0.232
Weight of Whole Flakes - Total	0.828	0.092	0.067

Specifically, because Factor 1 clustered on variables related to the total values of the selected metrics and Factor 2 clustered on variables related to the metrics' average values, these components appear related to the Quantity and Quality, respectively, of lithics produced by participants. Given that reduction intensity was the only variable loading that passed a threshold of more than |0.4|, the established minimum for factor stability (Guadagnoli and Velicer 1988), Factor 3 appears to represent an aspect of reduction intensity not captured by Factor 1. Because Factor 3 did not load on variables related to the number or weight of whole flakes or detached pieces, which should comprise the majority of core reduction, it would follow that Factor 3 accounts for the reduction of cores via shatter, or debitage that was too small (i.e. max dimension < 20mm) to measure.

To test this relationship, I calculated the percentage of mass removed from each core via shatter, using the following equation: $\left(\frac{Total\ Mass\ Removed - Detached\ Piece\ Mass\ Removed}{Total\ Mass\ Removed}\right) * 100.$ I then correlated this new variable with the cores' Factor 3 scores. The value of this correlation is small ($r = 0.245$) but positive and reports at a high level of significance ($p < 0.001$). Due to Factor 3's negative loading, higher factor values indicate that a participant reduced their cores less completely. In adults, Factor 3 highly correlated ($p < 0.001$) with the percentage of core mass removed via shatter, such that participants with higher factor scores also produced more shatter per core. This correlation is not sustained in child participants ($p = 0.057$), indicating that although children produced less shatter per core, their generally higher Factor 3 scores suggest they were not able to convert lower shatter generation into better core reduction and flake production. As such, Factor 3 appears to be a measure of Flaking Inefficiency, with lower scores indicating more efficient knapping and higher scores indicating less efficient knapping.

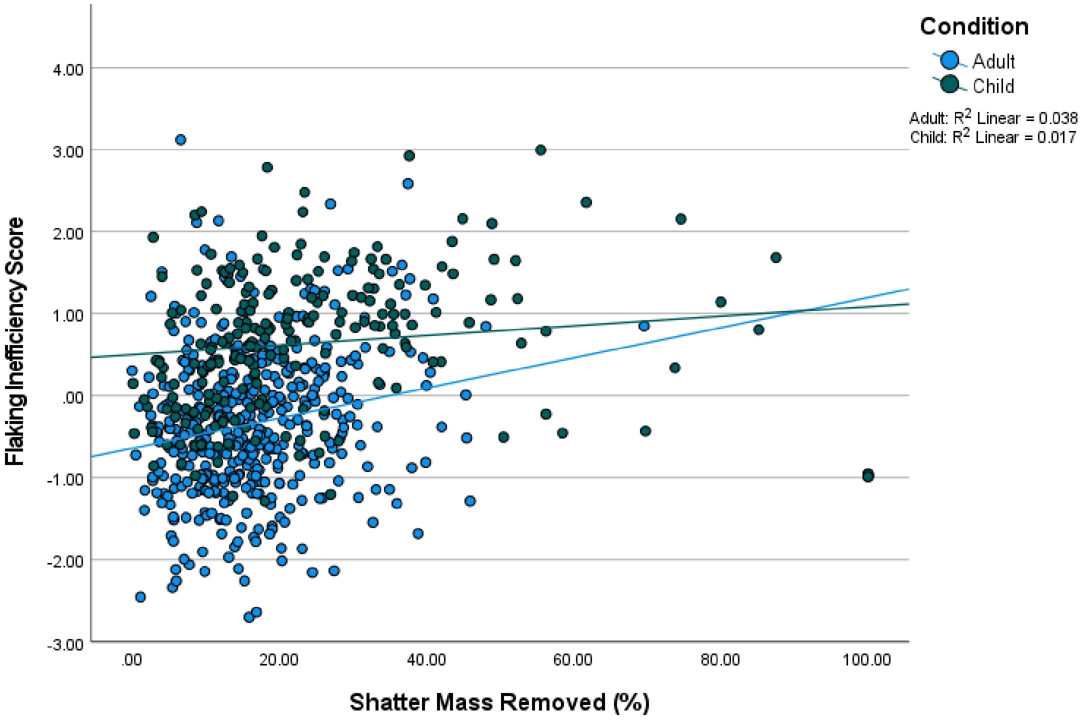


Figure 2.1: Flaking (In)Efficiency Score by Shatter Mass Removed. Each individual data point represents a participant core, and colors represent experimental condition.

Group-Level Differences of Knapper Skill

Using two-way ANOVAs, I sought to determine whether there were group-level differences in knapper performance of the above measures of skill (Quantity, Quality, and Flaking Inefficiency) at the developmental level (i.e. adults and children), between participants of different sexes (i.e. male and female), and across participant subgroups of age and sex (i.e. adult males, adult females, child males, and child females; hereafter interchangeable with “men,” “women,” “boys,” and “girls,” respectively). Of the 49 study participants, one adult identified as nonbinary and was excluded from all analyses based on participant sex.

There was no interaction of participant sex and condition for lithic Quantity (F(1,44) = 2.459, p = 0.124) (*Figure 2.2*) or Flaking Inefficiency (F(1,44) = 0.001, p = 0.977) (*Figure 2.4*). An interaction was present, however, for lithic Quality (F(1,44) = 5.291, p = 0.026) (*Figure 2.3*), such that adult male participants produced significantly higher Quality lithics than participants from all other subgroups (p = 0.010). This result may be accounted for by a notable variation in participant grip strength, discussed below in greater detail (§ Grip Strength), wherein adult men demonstrated grip strength values that were, on average, much larger than all other participant subgroups (i.e. adult women and children of both sexes), though adult women still demonstrated grip strength scores that were statistically larger than those of children (p < 0.001). Given the interaction between participant age (i.e. condition) and sex for lithic Quality, I removed adult male participants from the sample and conducted an independent samples t-test comparing the adult female participants' Quality scores with those of children of both sexes. The result of this test showed that there was no statistical difference between these groups (p = 0.071).

Although there was no interaction between participant sex and condition for lithic Quantity and Flaking Inefficiency, both of these knapper skill metrics demonstrated main effects results along these participant categories. For condition, there were significant main effect group level differences between adults and children of Quantity (p < 0.001, Cohen's d = 0.846) and Flaking Inefficiency (p < 0.001, Cohen's d = -1.094). Adults produced more lithics (*Figure 2.2*) of higher quality (*Figure 2.3*). Children, on average, demonstrated a higher Flaking Inefficiency scores than adults (*Figure 2.4*), which, due to the factor's negative loading, indicates that children removed less mass from their cores than adults and, overall, knapped their cores less effectively (*Figure 2.1*).

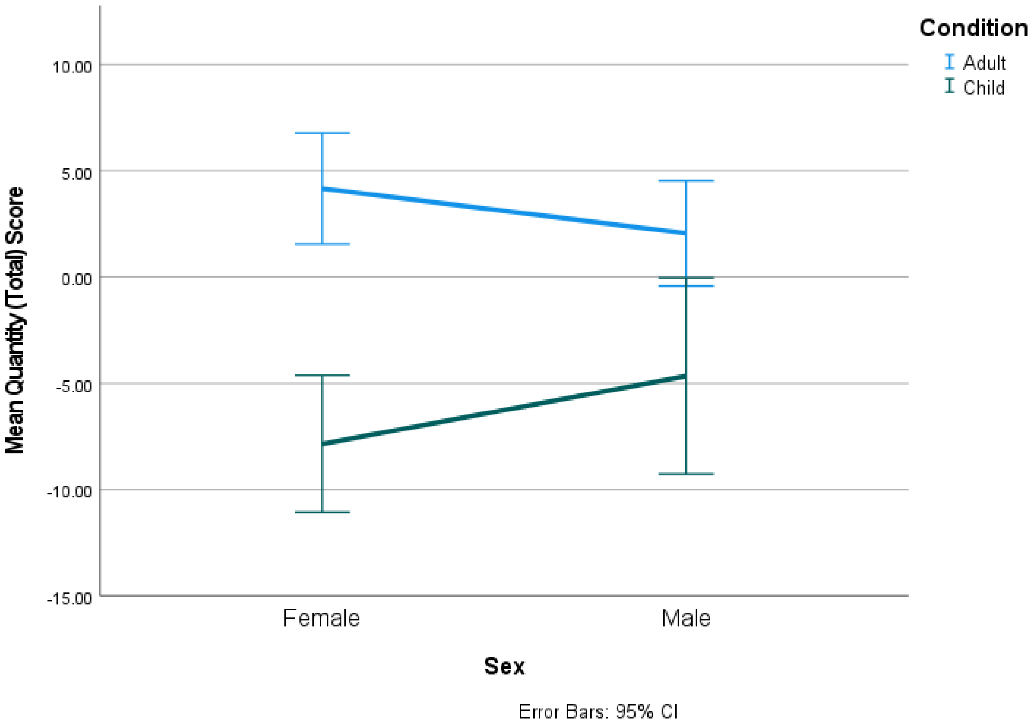


Figure 2.2: Mean Lithic Quantity by Participant Sex and Condition. On average, adults produced a greater total Quantity of lithics than children, and female participants produced more than male participants.

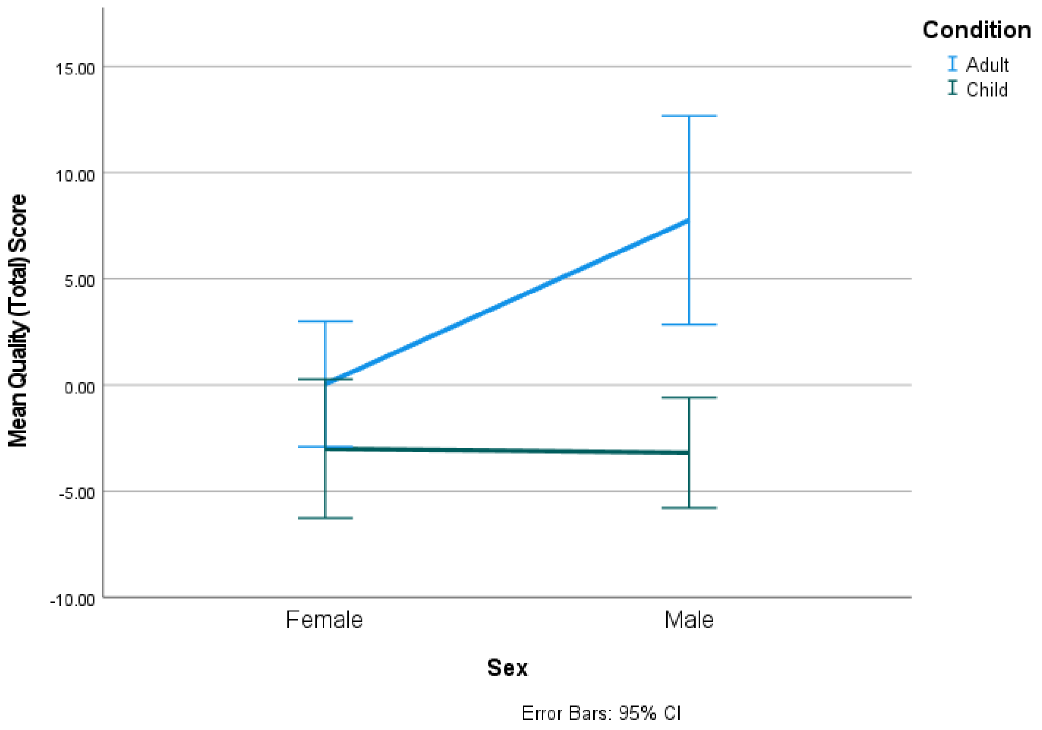


Figure 2.3: Mean Lithic Quality by Participant Sex and Condition. On average, adults produced lithics of better Quality than children. Adult male participants produced better Quality lithics than all other participant subgroups. Adult female participants produced lithics of a Quality that were statistically indistinguishable from those of children.

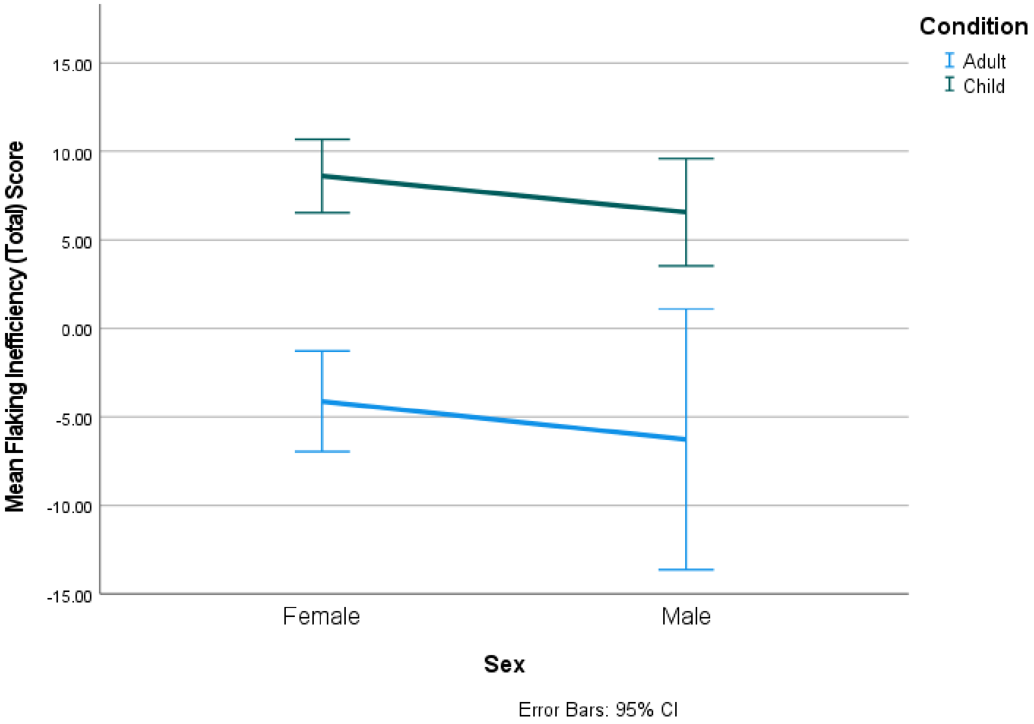


Figure 2.4: Mean Flaking Inefficiency by Participant Sex and Condition. Children less intensely reduced their cores than adults and were more likely to produce shatter, as opposed to useable detached pieces. There is no significant difference in Flaking Inefficiency between male and female participants.

Male and Female participants demonstrated group-level differences of Quantity (p = 0.005) (*Figure 2.2*) but not for Flaking Inefficiency (p = 0.075) (*Figure 2.4*), with female participants producing slightly more lithics. The effect sizes of these findings are relatively small (for Quantity, Cohen's d = -0.228; for Quality, Cohen's d = 0.176), indicating that participant sex, by itself, may not be the most robust predictor of knapper outcomes. Within male participants, lithic Quantity positively correlated with Quality (r(17) = 0.496, p = 0.031) and demonstrated a marginal negative correlation with Flaking Inefficiency (r(17) = -0.441, p – 0.059), and Quality and Flaking Inefficiency negatively correlated with each other (r(17) = -

0.480, $p = 0.038$). Among female participants, Quality negatively correlated with Flaking Inefficiency ($r(27) = -0.497$, $p = 0.006$).

The gulf between the lithic Quality of adult male and adult female participants highlighted the necessity of interrogating the data for any further discrepancies between participants of different sexes within conditions, thus creating four distinct participant subgroups: adult males ($n = 8$, “men”), adult females ($n = 21$, “women”), child males ($n = 11$, “boys”), and child females ($n = 8$, “girls”). Individually, these groups are small and their sizes uneven, so any results reported here on these separate groups should be taken as preliminary.

Using independent samples t-testing, I identified a statistically significant difference ($p < 0.001$) of Quality between adult male and adult female participants, as alluded to above, with male participants producing lithics of better Quality than female participants (*Figure 2.3*). This difference was not maintained for lithic Quantity or Flaking Inefficiency. Among child participants, male children were more likely to produce more lithics (Quantity: $p = 0.010$) (*Figure 2.2*), with female children reducing their cores less intensely, overall (Flaking Inefficiency: $p = 0.015$) (*Figure 2.4*). There was no statistical difference of lithic Quality among child participants.

Learning Effects

To determine whether participants improved over the course of their five days of practice in their lithic Quantity and Quality output or experienced changes in their ability to remove mass from the core in the form of useable detached pieces, as opposed to shatter (Flaking Inefficiency), I conducted a series of one-way ANOVA analyses with Day as the independent variable and the respective participant group as the dependent variable. When all participants were tested together, a learning effect in Quantity was demonstrated (F = 8.834, p < 0.001). Tested separately, this learning effect in Quantity remained in children (F = 7.197, p < 0.001), adults (F = 4.053, p = 0.003), male participants (F = 5.253, p < 0.001), and female participants (F = 4.068, p = 0.003). Interestingly, when participants were tested by subgroup, the learning effect for Quantity remained in all groups (Adult Females: F = 3.062, p = 0.017; Child Males: F = 4.323, p = 0.003; Child Females: F = 3.961, p = 0.006) except for adult males (F = 1.536, p = 0.197). In each statistically significant case, the Quantity of lithics participants produced increased from Day 1 to Day 5 (*Figure 2.5*).

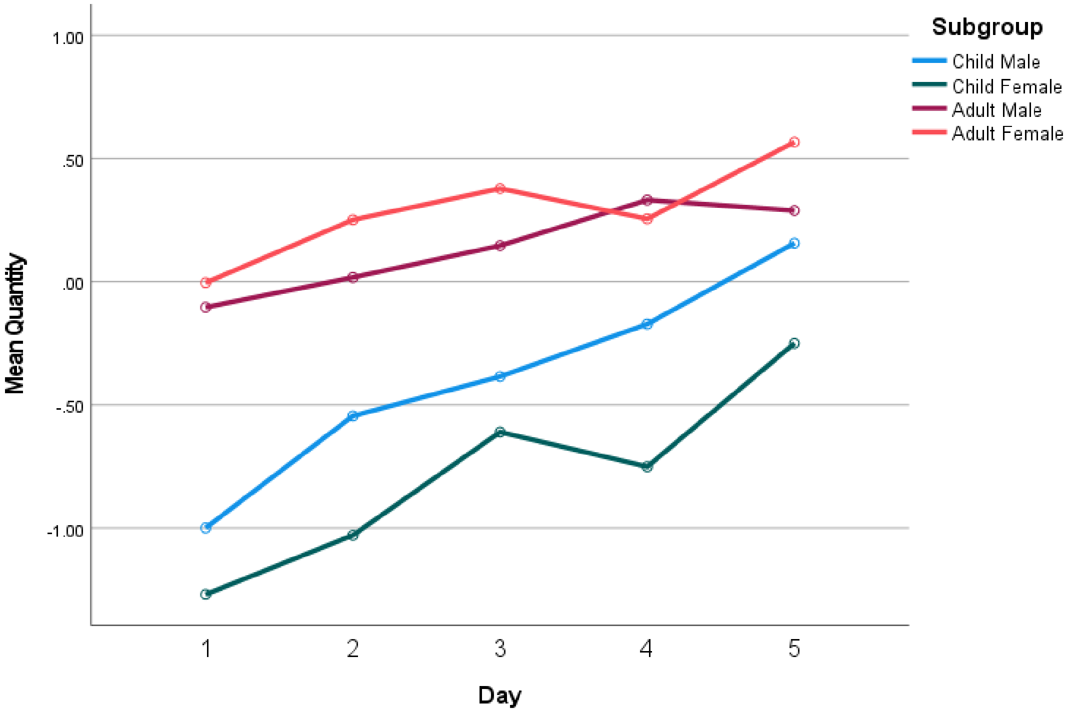


Figure 2.5: Learning Effects of Quantity by Subgroup. All subgroups, except for Adult Males, increased their lithic output over the course of the five days of knapping practice.

No learning effects were demonstrated at any level of analysis for lithic Quality (*Figure 2.6*) or Flaking Inefficiency (*Figure 2.7*). However, despite the results not meeting the threshold of statistical significance, it is worth noting that for all participant groupings, the Quality of lithic materials appeared to decrease over the five days of knapping practice. Because the training time was limited to five days, 30-minutes per day, for a total of 2.5 hours of practice, it is possible that participants were entering into a new phase of their knapping knowledge at the end of the experimental period – one where they were beginning to better understand the mechanics of the task without also having acquired the requisite motor skills, gained through long hours of deliberate practice, to achieve expert-level performance. Should the experiment have continued at the same rate of practice over a longer period of time, one would expect participant performance to follow the Power Law of Practice, similar to that demonstrated in Pargeter, Khreisheh, and Stout (2019).

;

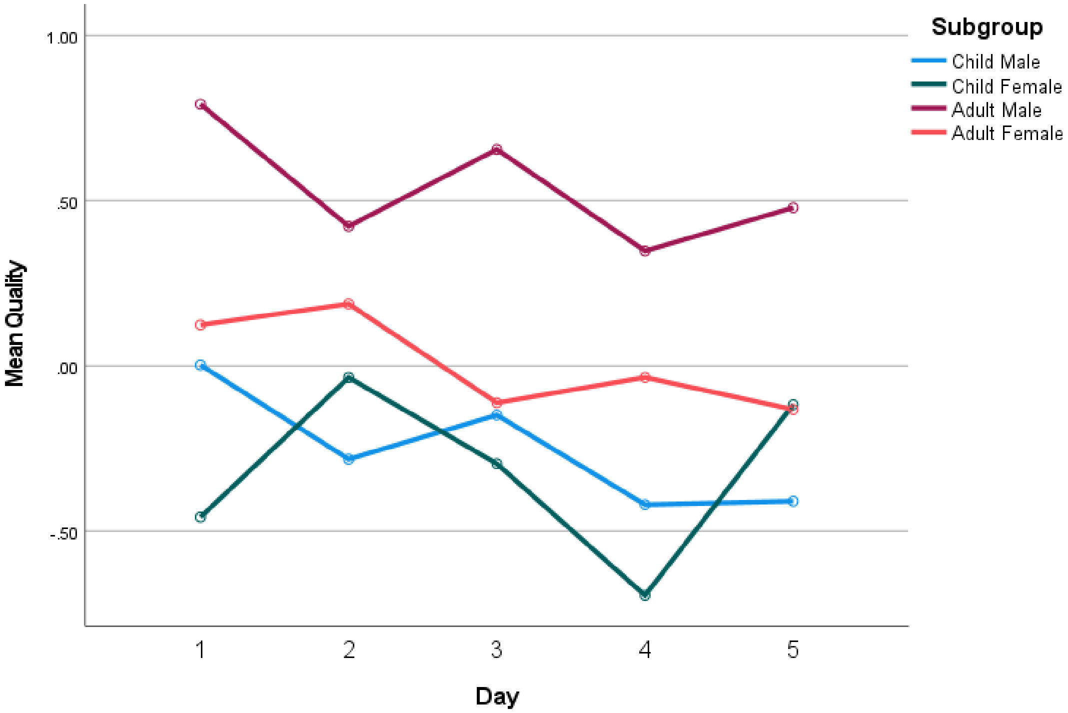


Figure 2.6: Learning Effects of Quality by Subgroup. None of the subgroups reported significant learning effects of lithic Quality.

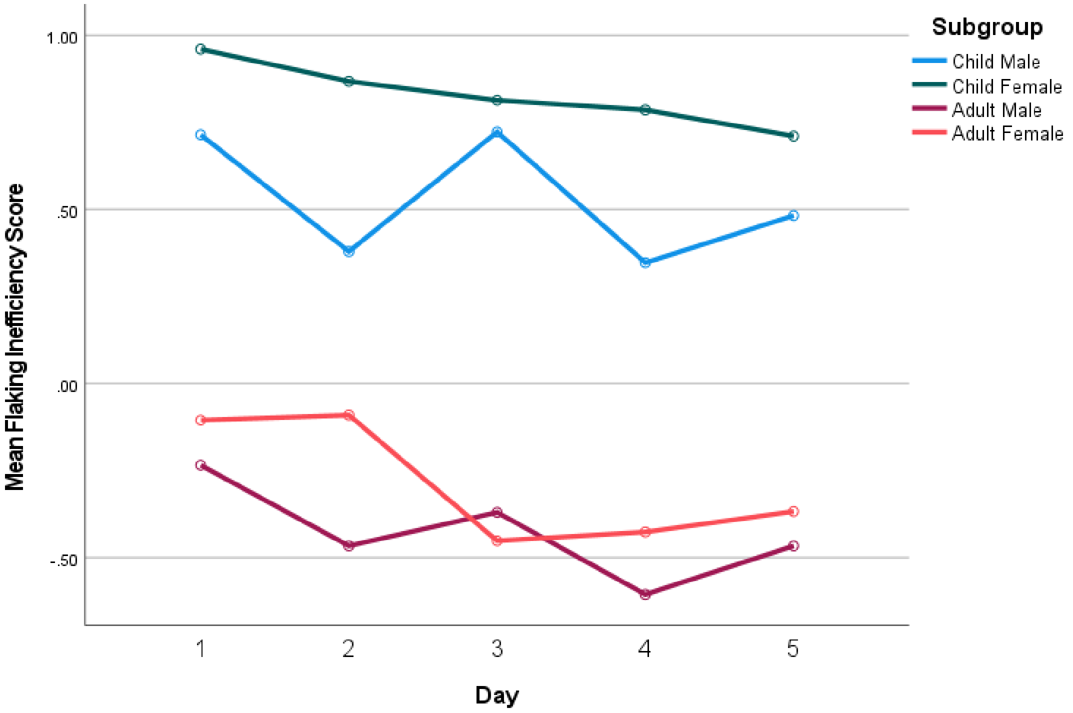


Figure 2.7: Learning Effects of Flaking Inefficiency by Subgroup. None of the subgroups reported significant learning effects of Flaking Inefficiency.

To test for individual differences in learning, I calculated the average Quantity, Quality, and Flaking Inefficiency scores for each participant on Days 1 and 5 of the experiment and subtracted the former from the latter to create a delta value. These values were then correlated with each participant's psychometric and motor testing scores. When all subjects were tested together, none of the skill metric and psychometric/motor pairings yielded significant results. In adults, there were negative learning effect correlations of lithic ΔQuantity with Fitts' Law Test response times (r(27) = -0.423, p = 0.022) (*Figure 2.8*) and lithic ΔQuality with MRT scores (r(27) = -0.376, p = 0.045) (*Figure 2.9*). Children demonstrated positive learning effect correlations of ΔQuantity with MRT score (r(17) = 0.552, p = 0.014) (*Figure 2.10*). and ΔQuality with the BFI: Agreeableness factor (r(17) = 0.575, p = 0.010) (*Figure 2.11*). When tested by participant sex, male participants reported a positive relationship between ΔQuality and Agreeableness (r(17) = 0.531, p = 0.019) (*Figure 2.12*), whereas female participants showed a negative correlation between ΔQuantity and Openness scores (r(26) = -0.432, p = 0.022) (*Figure 2.13*). Among participant subgroups, adult male participants demonstrated a negative correlation of Conscientiousness and ΔFlaking Inefficiency (r(6) = -0.715, p = 0.046) (*Figure 2.14*), and female children showed a positive correlation of Conscientiousness and ΔQuantity (r(6) = 0.716, p = 0.046) (*Figure 2.15*). Adult female participants reported negative correlations between Openness and Quantity (r(18) = -0.499, p = 0.025) (*Figure 2.16*), Agreeableness and ΔQuality (r(18) = -0.483, p = 0.031) (*Figure 2.17*), and Fitts' Law Test response time and ΔQuantity (r(18) = -0.508, p = 0.022) (*Figure 2.18*). Male children demonstrated positive a correlation between Agreeableness and ΔQuality (r(9) = 0.778, p = 0.005) (*Figure 2.19*) and a marginal negative correlation between MRT score and ΔQuality (r(9) = -0.587, p = 0.052) (*Figure 2.20*).

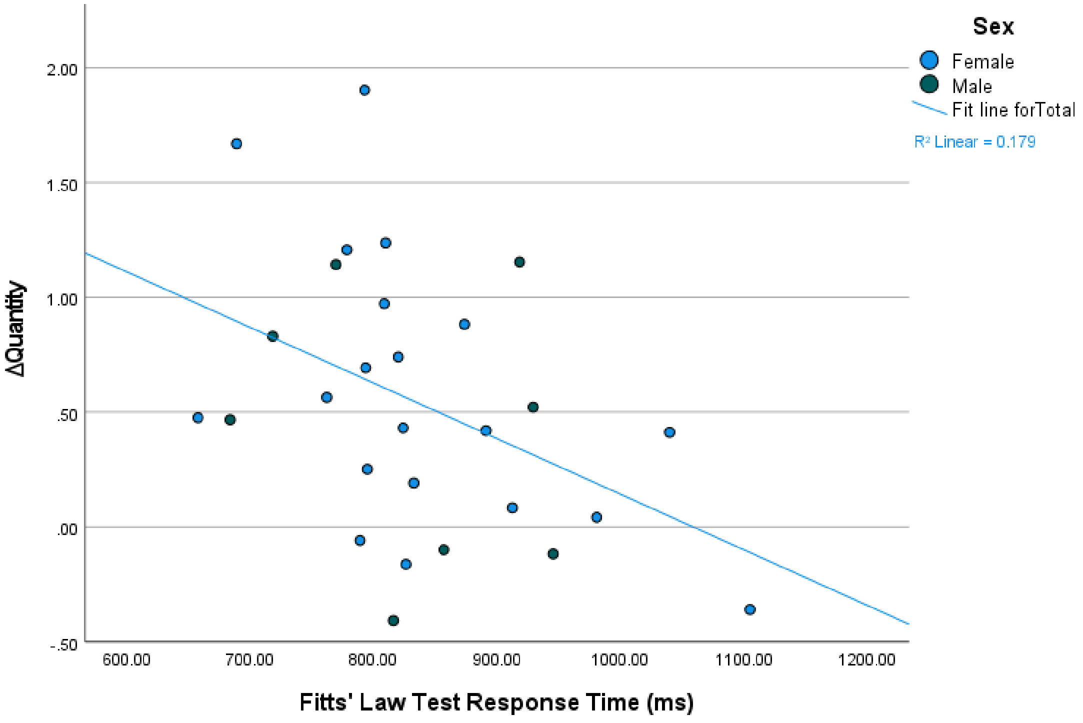


Figure 2.8: Effect of Fitts' Law Test Response Time on ΔQuantity in Adults. Adult participants with faster response times also saw improvements in their lithic Quantity over the course of the training period.

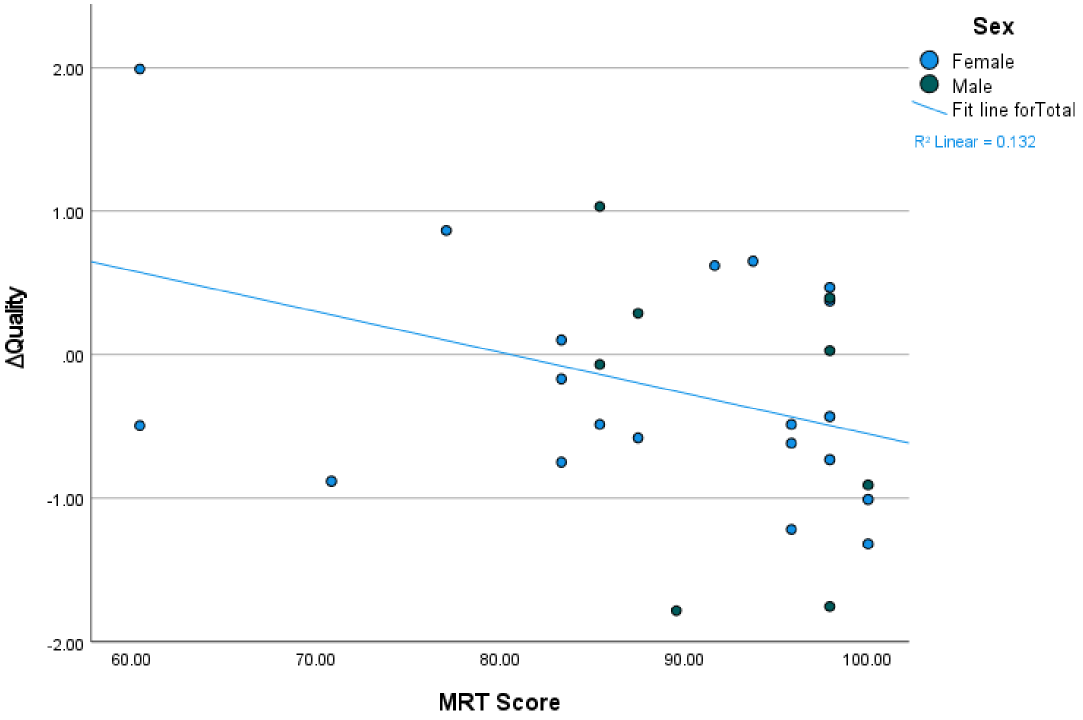


Figure 2.9: Effect of MRT Score on ΔQuality in Adults. Adult participants who performed better in the MRT task experienced a regression in lithic Quality over the course of the training period.

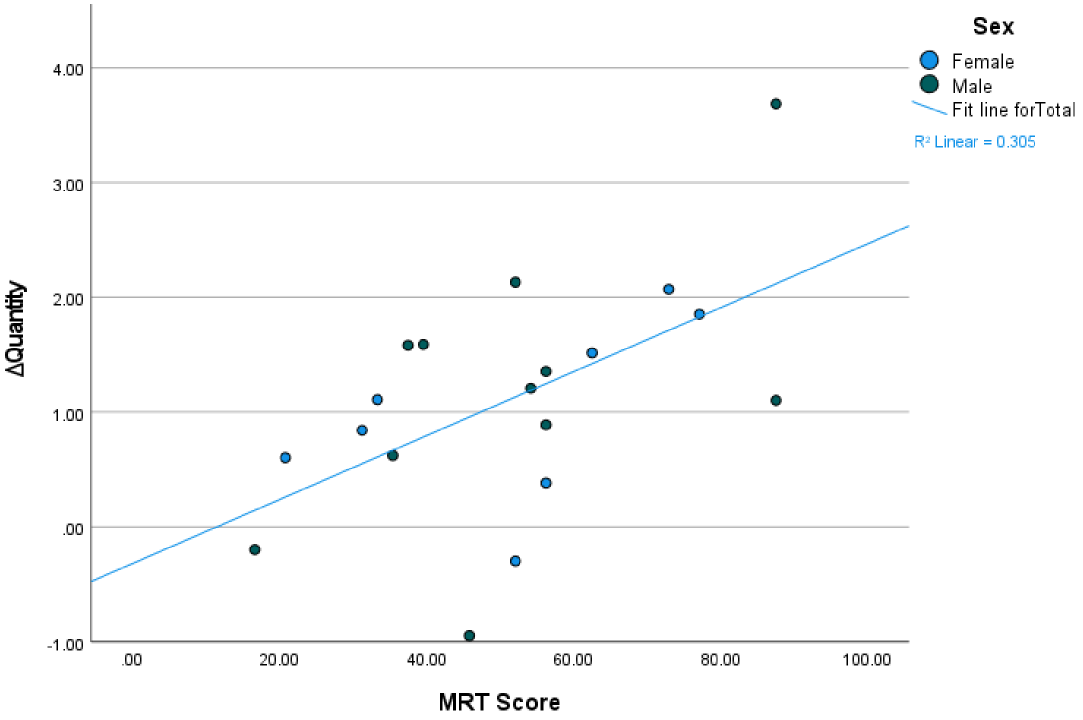


Figure 2.10: Effect of MRT Score on ΔQuantity in Children. Child participants with better MRT scores saw improvements in their lithic output.

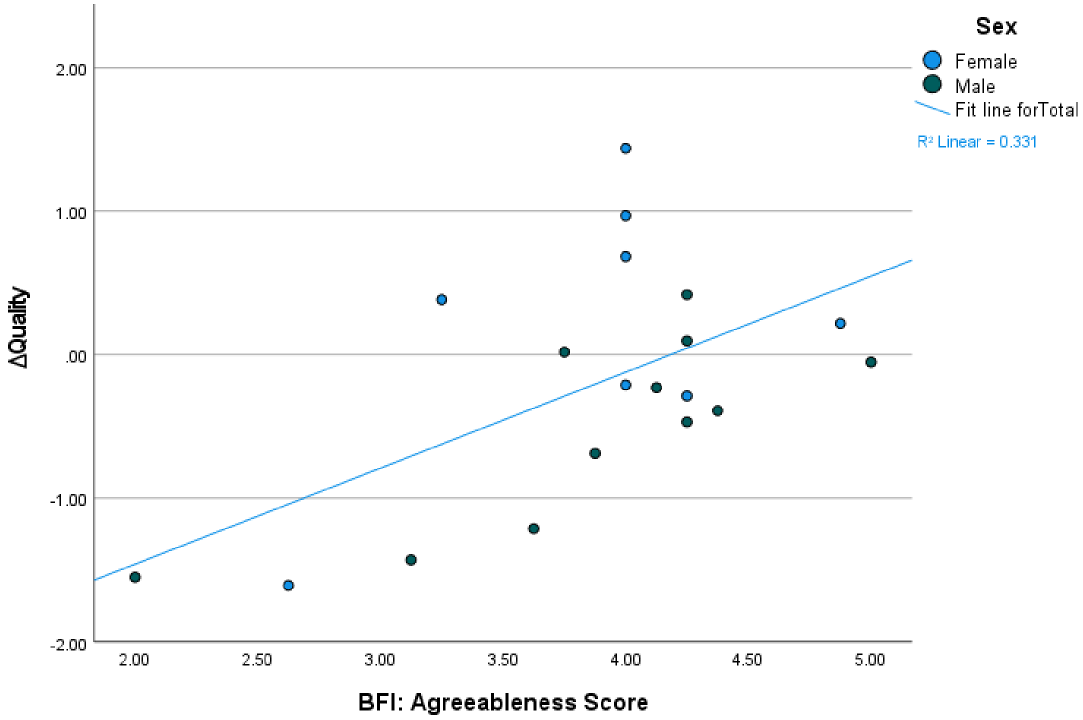


Figure 2.11: Effect of BFI: Agreeableness Score on ΔQuality in Children. Child participants who were more Agreeable also saw improvements in the Quality of their lithics over the training period.

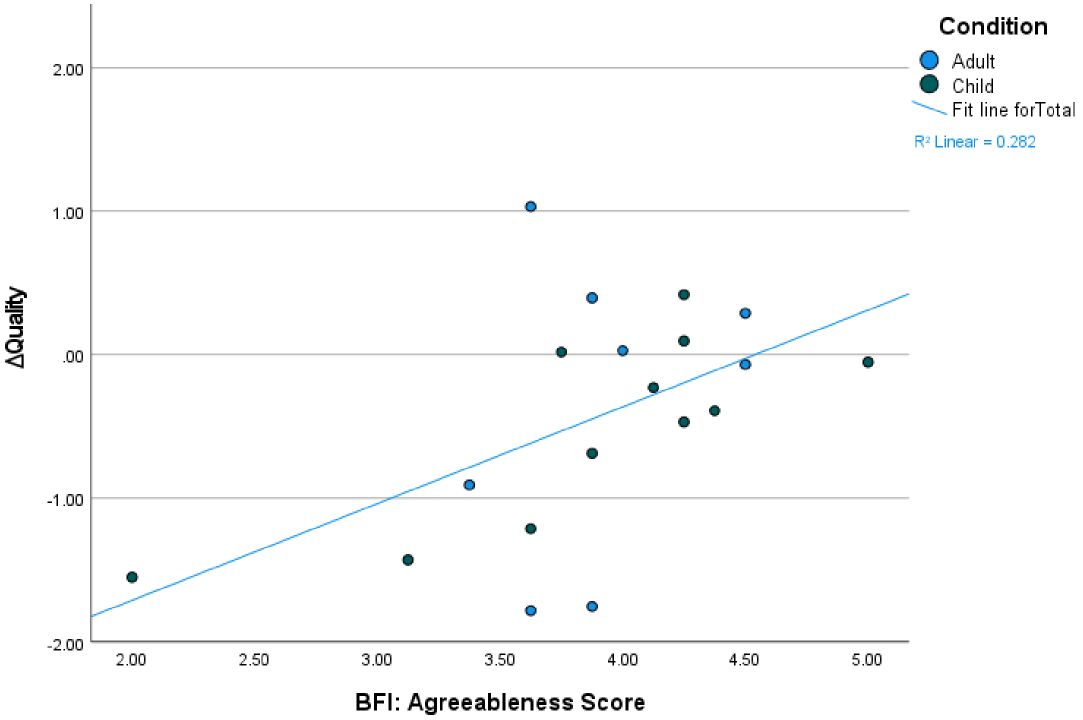


Figure 2.12: Effect of BFI: Agreeableness Score on ΔQuality in Male Participants. More Agreeable male participants saw improvements in lithic Quality over the training period.

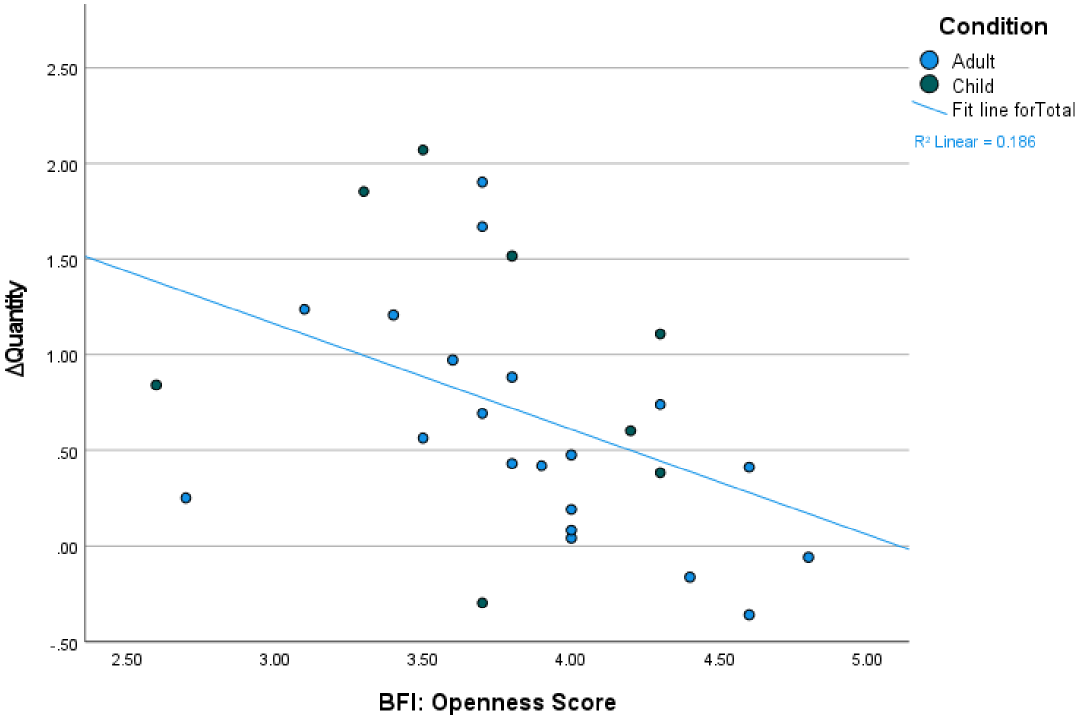


Figure 2.13: Effect of BFI: Openness Score on ΔQuantity in Female Participants. Female participants with higher Openness scores demonstrated a reduction in their lithic output over the training period.

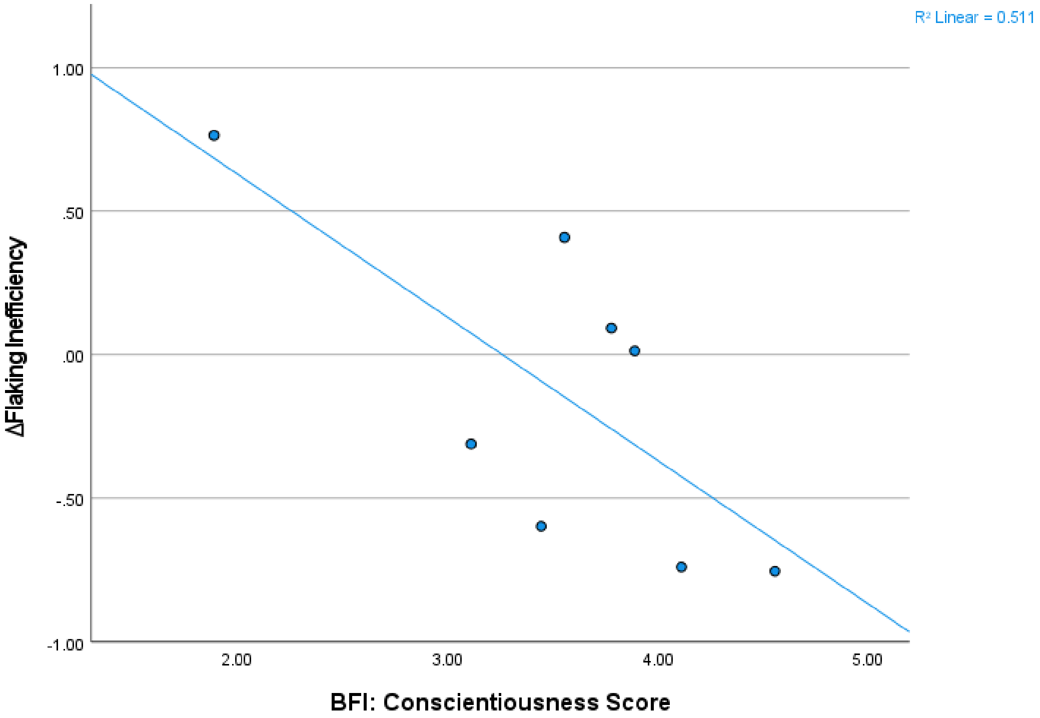


Figure 2.14 Effect of BFI: Conscientiousness Score on ΔFlaking Inefficiency in Adult Male Participants. More Conscientious male participants saw improvements to their Flaking Efficiency over the course of the training period.

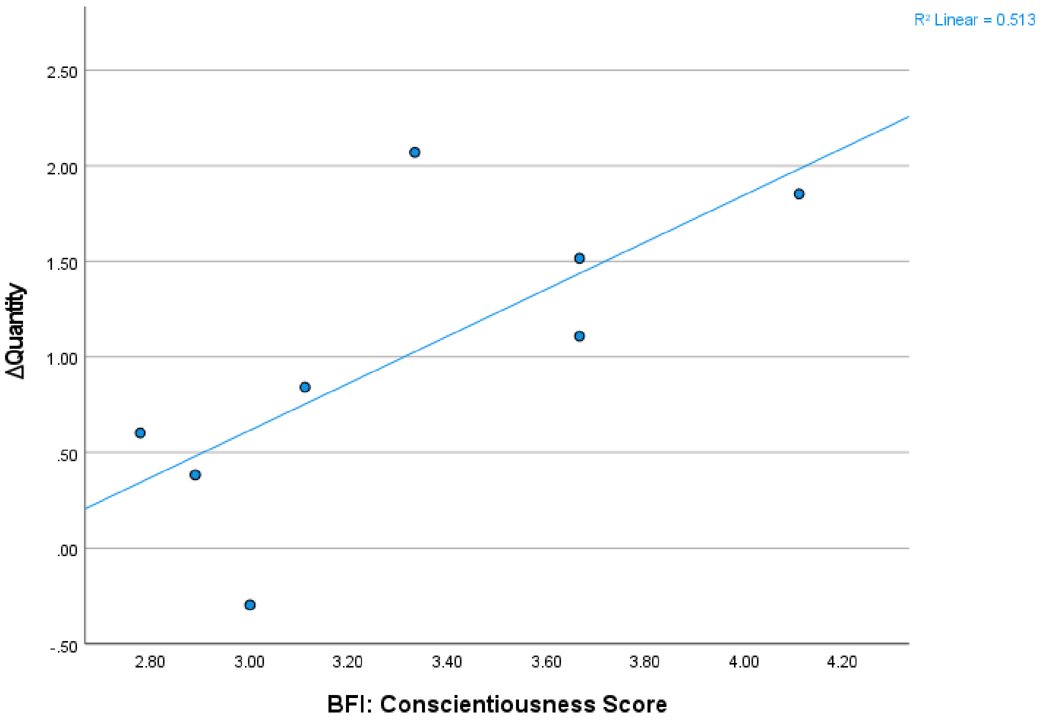


Figure 2.15 Effect of BFI: Conscientiousness Score on ΔQuantity in Child Female Participants. More Conscientious female children were also more likely to see improvements in their lithic output over the training period.

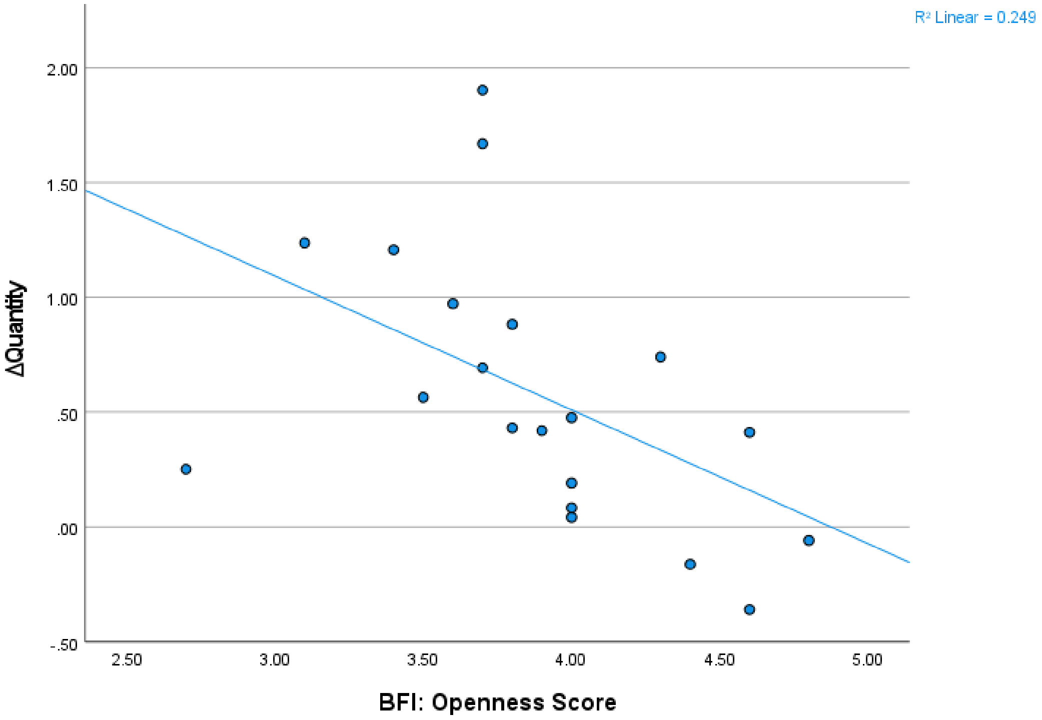


Figure 2.16: Effect of BFI: Openness Score on ΔQuantity in Adult Female Participants. Adult female participants with higher Openness scores saw a reduction in their lithic output over the training period.

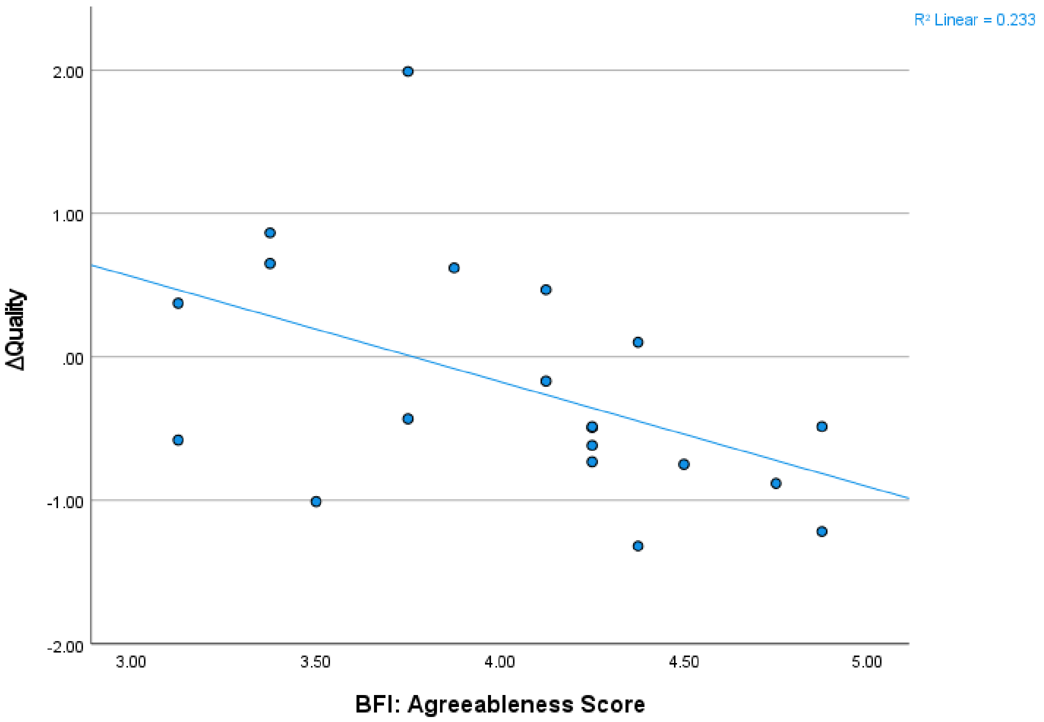


Figure 2.17: Effect of BFI: Agreeableness Score on ΔQuality in Adult Female Participants. More Agreeable adult female participants experienced a reduction in their lithic Quality over the training period.

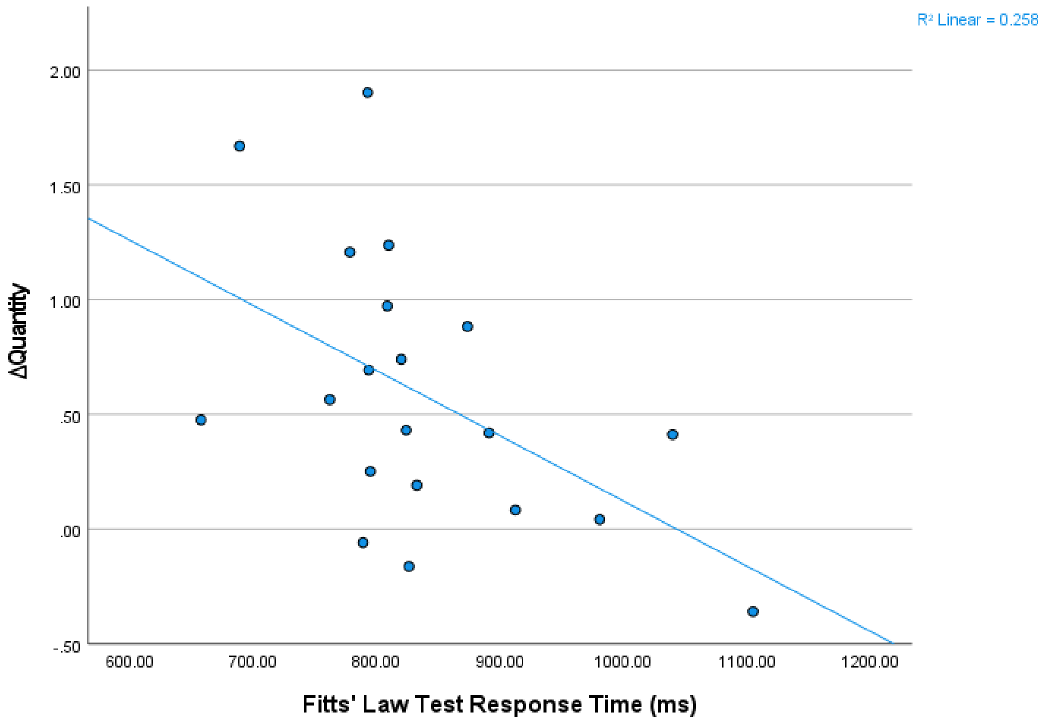


Figure 2.18: Effect of Fitts' Law Test Response Time on ΔQuantity in Adult Female Participants. Adult female participants with faster FLT response times were also more likely to produce fewer lithics at the end of the training period.

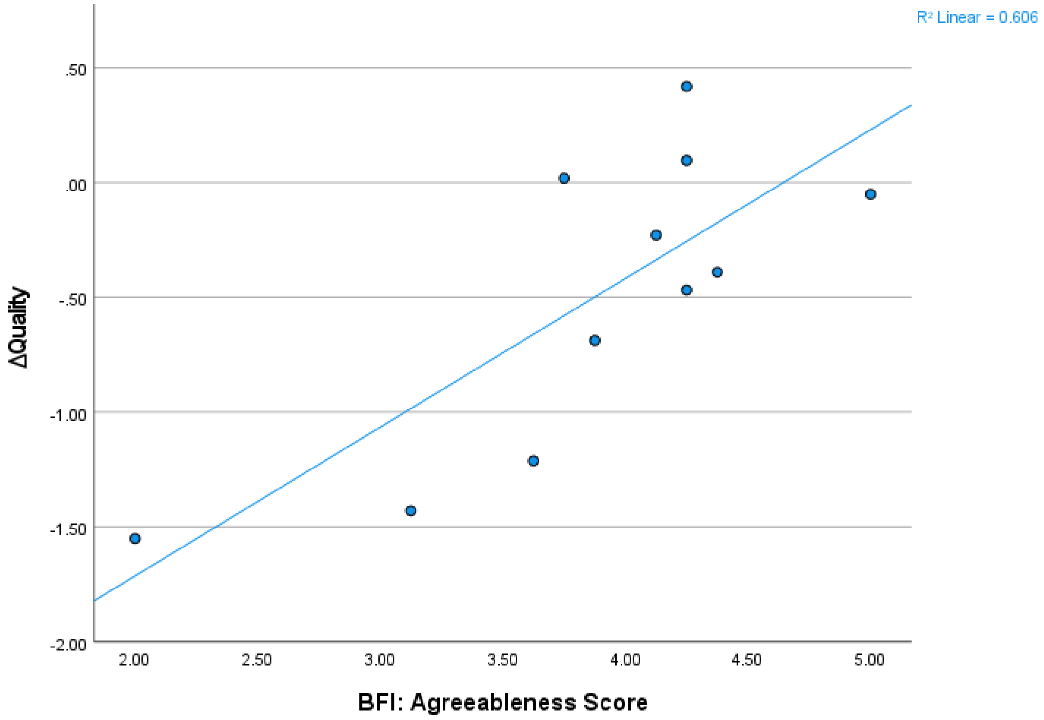


Figure 2.19: Effect of BFI: Agreeableness Score on ΔQuality in Child Male Participants. Male children who were more Agreeable were also more likely to experience an increase in lithic Quality over the training period.

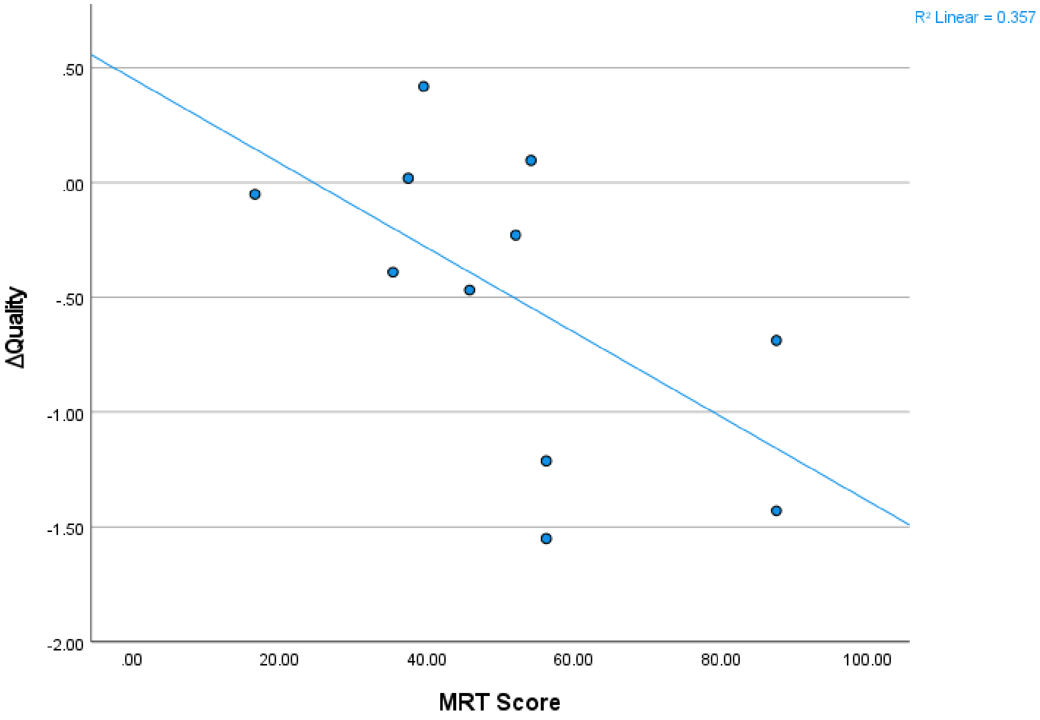


Figure 2.20: Effect of MRT Score on ΔQuality in Child Male Participants. Male children with higher MRT scores saw a decrease in the Quality of their lithic products over the course of the training period.

Core Size and Shape

Because core shape was not standardized, I wanted to examine whether the random sampling of cores was unbiased across the two experimental conditions (i.e. adults and children). An independent samples t-test of Factors 1 ("Flatness") and 2 ("Roundness") calculated from the above mentioned PCA of core shape revealed no difference between adult's and children's cores along Factor 1 ($p = 0.082$); however, there was a significant difference between them along Factor 2 ($p = 0.011$). To further investigate which aspects of core shape may be contributing to this difference, I ran an additional independent samples t-test on the cores' scale-free linear dimensions. Although there was no significant difference in length between the cores assigned to children and those assigned to adults (p = 0.808), there were differences in core width (p = 0.018) and thickness (p = 0.011). It should be noted, however, that there was no clear bimodal distribution of core shapes, as demonstrated in *Figure 2.21*.

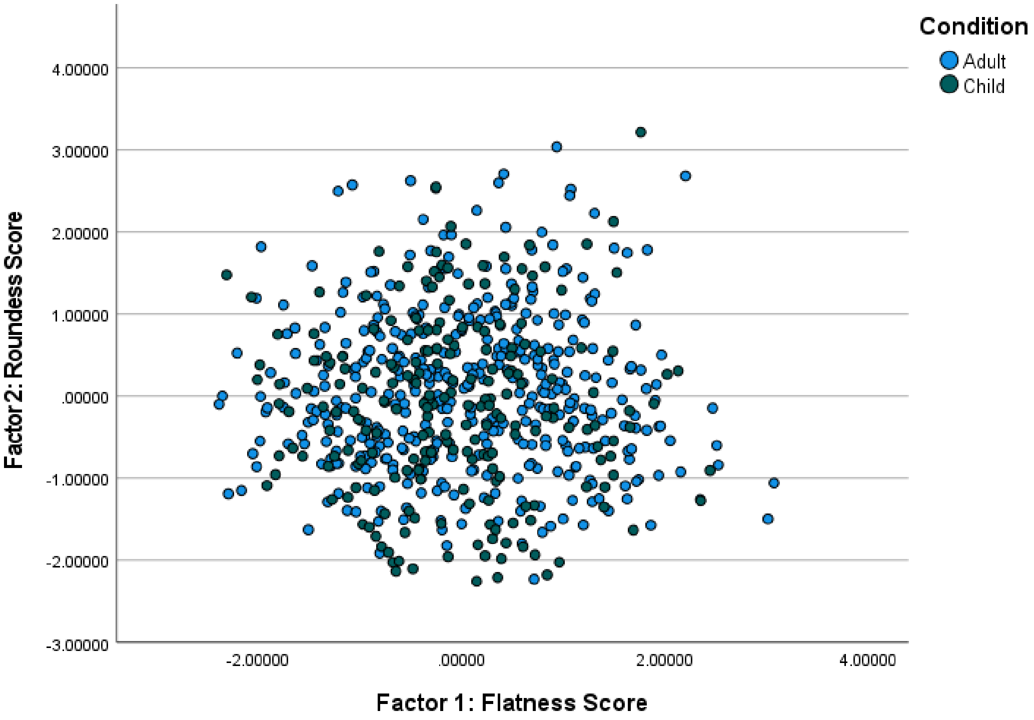


Figure 2.21: Distribution of Core Shapes Across Experimental Conditions. Each data point represents a core, and there is no apparent bimodal distribution of core shapes between conditions.

Because these significant results were not apparent when graphed, I looked to Cohen's d to examine the effect size of these findings. According to Cohen (1962), a Cohen's d value of less than 0.25 is considered to be indicative of a small effect size. The values for core width and thickness were $d = 0.197$ and $d = 0.212$, respectively, and because of this, it is likely that the impact of core shape on knapper output was minimal.

To examine the degree and nature of the influence of core shape on knapping outcome, I ran linear regressions of each shape score (i.e. Factor 1: Flatness and Factor 2: Roundness) against measures of knapper skill (i.e. Quantity, Quality, and Flaking Inefficiency). When all participants were tested together, both core Flatness and Roundness appeared to impact the Quantity (Flatness: $p < 0.001$; Roundness: $p < 0.001$) of knapper output, but not the Quality (Flatness: $p = 0.418$; Roundness: $p = 0.175$) of lithic products or Flaking Inefficiency of cores (Flatness: $p = 0.420$; Roundness: $p = 0.387$). This trend remained in adult participants (Quantity: Flatness: $p < 0.001$ and Roundness: $p < 0.001$; Quality: Flatness: $p = 0.673$ and Roundness: $p = 0.457$; Flaking Inefficiency: Flatness: $p = 0.095$ and Roundness: $p = 0.621$). For children, core Flatness ($p < 0.001$), but not Roundness ($p = 0.075$), impacted Quantity, a trend that was also demonstrated with Flaking Inefficiency (Flatness: $p = 0.007$; Roundness: $p = 0.225$). Neither shape factor impacted lithic Quality within child participants (Flatness: $p = 0.806$; Roundness: $p = 0.622$).

That shape should influence the Quantity of lithic products made from a core is not surprising, given that, especially in novices who are less adept at flexibly adapting to core morphological affordances, the availability of viable platform angles on the starting core will impact the success of both initial and successive flake removals. This being said, “flatter” cores, which typically have a greater number of starting viable platform angles due to their comparatively longer lengths and widths and shorter thicknesses, should yield higher Quantity values than their more difficult to knap “rounder” counterparts. When all participants were tested together in an independent samples t-test, there was a significant difference in Quantity between "flatter" and "rounder" cores ($p = 0.007$). This trend persisted in adults ($p = 0.009$) but disappeared in children ($p = 0.072$), thus indicating that core shape was a factor in knapper output for adults but not for children. Interestingly, however, adults, on average, received cores that were wider (Scale-Free Mean Width: a. for Adults: 1.03; b. for Children: 0.99) and thicker (Scale-Free Mean Thickness: a. for Adults: 1.05; b. for Children 0.99) than those children received, which should have placed them at a disadvantage for their Quantity scores. Converse to this prediction, adults greatly outperformed children in Quantity ($p < 0.001$) at a large effect size (Cohen's $d = 0.846$), which suggests both that starting core shape did not hinder adult knapping performance and that the effect of condition on Quantity may be even greater than reported here had core shape been standardized between adult and child participants.

Psychometric and Motor Testing

Due to developmental differences of motor and cognitive ability, we predicted that adults' and children's respective performance on a series of psychometric and motor testing related to upper body strength, personality, and executive functioning may be predictive of knapping outcomes. All analyses conducted to determine these relationships were performed as linear regressions, with the given psychometric test (i.e. grip strength, MRT, Fitts' Law Test, Stroop Raw Color-Word Test, or BFI) acting as the independent variable and the knapping skill measure (i.e. Quantity, Quality, or Flaking Inefficiency) acting as the dependent variable, within each testing group. Group-based differences were determined using one-way ANOVAs.

Grip Strength

Group-Based Differences

As one may expect, adults, on average, demonstrated higher grip strength scores than children ($p < 0.001$). There was no statistically significant difference in grip strength between male and female participants, which may be attributable to the fact that children do not develop sex-based differences in upper body strength until adolescence (Neu et al. 2002; Wind et al. 2010). Within subgroups (*Figure 2.22)*, adult male participants outperformed all other subgroups at a significance level of $p < 0.001$. Adult female participants were, on average, stronger than children of both sexes, at $p = 0.007$, when compared with female children, and at $p = 0.009$, when compared with male children. Consistent with the literature, male and female children did not share demonstrate a difference in grip strength; however, within child participants, grip strength did significantly increase with age ($p = 0.001$) (*Figure 2.23*). As expected, this trend continued across developmental groups ($p = 0.003$) (*Figure 2.24).*

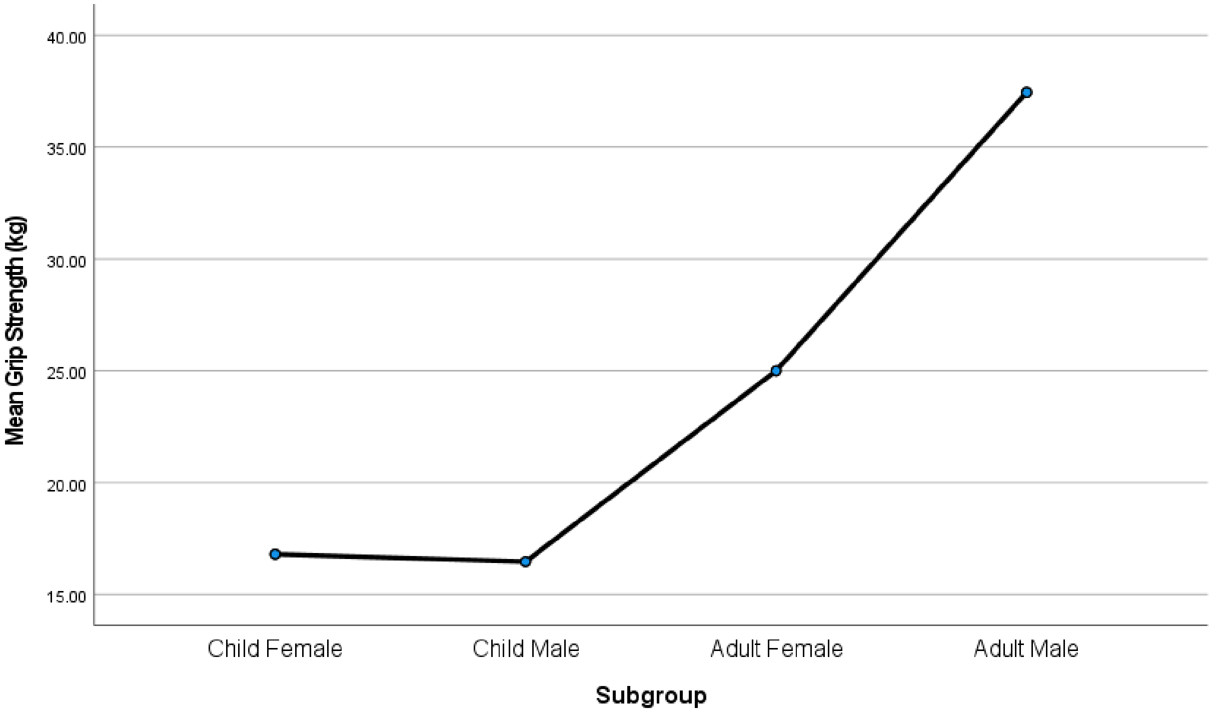


Figure 2.22: Mean Grip Strength Across Experimental Subgroups. Adult male participants significantly outperformed all other subgroups ($p < 0.001$), and adult female participants demonstrated greater grip strength values than children of both sexes ($p = 0.007$ and $p = 0.009$ when compared with female children and male children, respectively. Children did not show significant sex-based differences of grip strength.

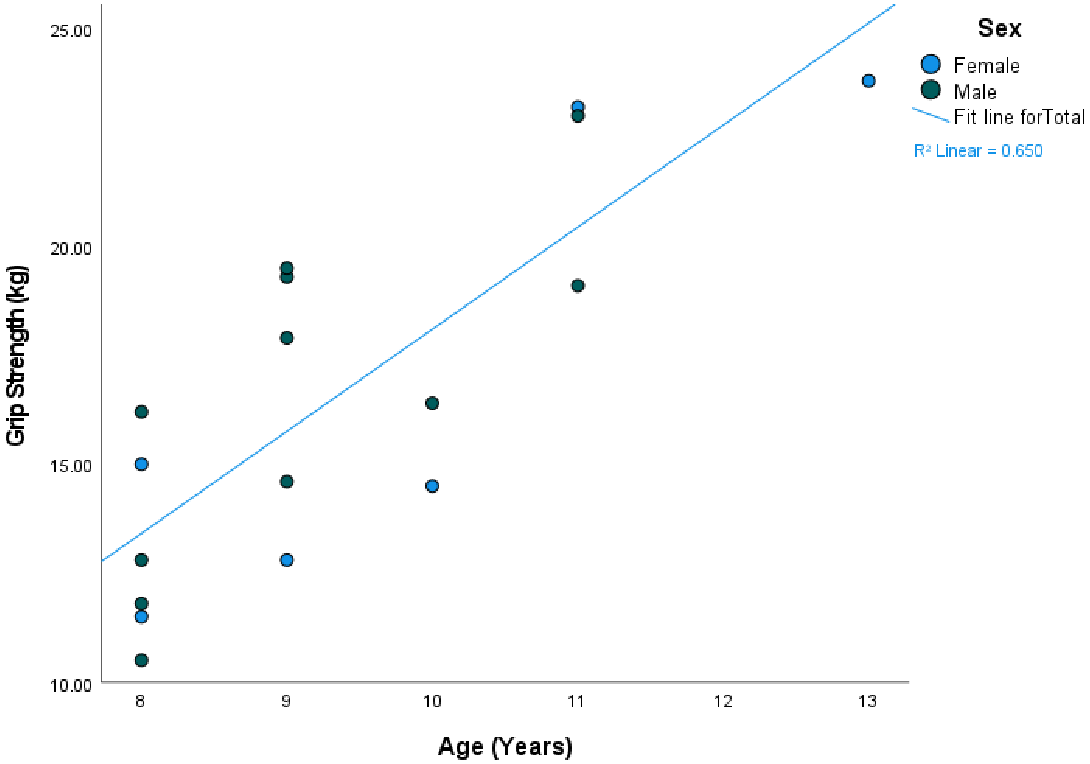


Figure 2.23: Effect of Age on Grip Strength in Children. Although children did not demonstrate sex-based differences of grip strength, they did show a positive relationship of grip strength with age.

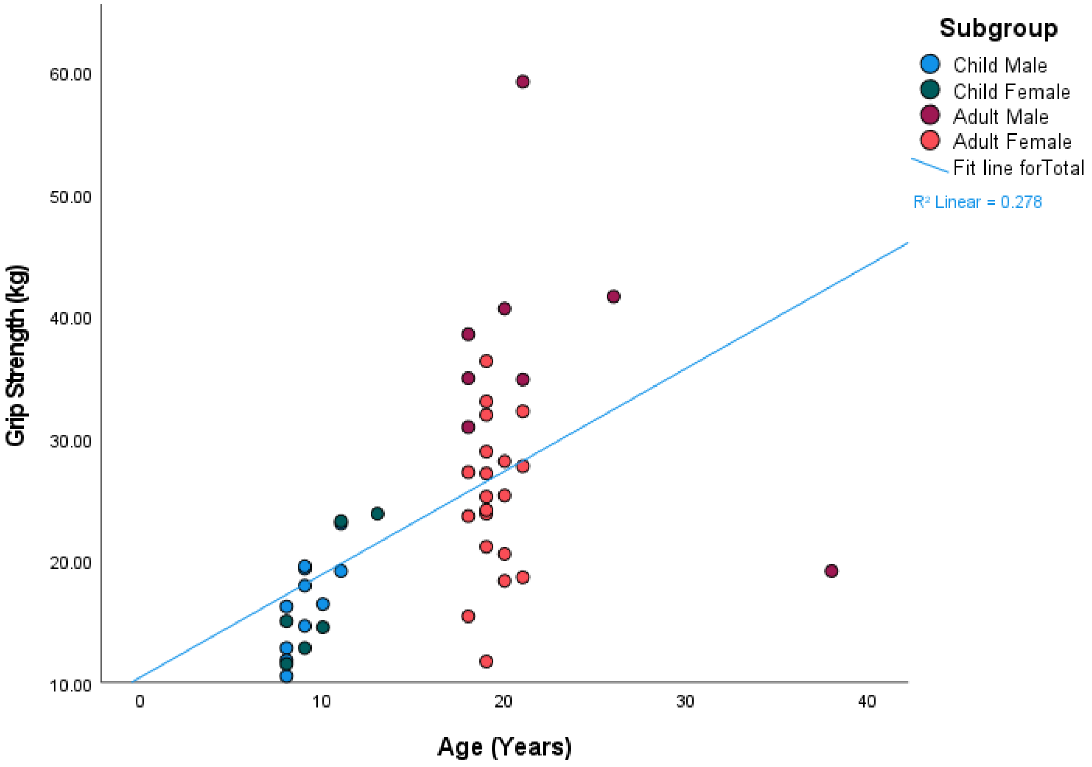


Figure 2.24: Effect of Age on Grip Strength in All Participants. Grip strength, generally, increases with age across experimental conditions.

Correlation with Knapper Skill

When all participants were tested together, grip strength was positively predictive of lithic Quantity (r(44) = 0.405, p = 0.005) (*Figure 2.125*) and Quality (r(44) = 0.487, p < 0.001) (*Figure 2.26*) and negatively predictive of Flaking Inefficiency (r(44) = -0.392, p = 0.007) (*Figure 2.27*). Tested by condition, a significant positive correlation remained in child participants for lithic Quality (r(15) = 0.561, p = 0.019), but not Quantity or Flaking Inefficiency, and all correlations disappeared entirely from adult participants. Tested by participant sex, male participants demonstrated positive correlations of grip strength with lithic Quality (r(17) = 0.725, p < 0.001) and marginally of Quantity (r(17) = 0.453, p = 0.052) and a negative correlation with

Flaking Inefficiency (r(17) = -0.551, p = 0.015). In female participants, grip strength was positively predictive of lithic Quantity (r(24) = 0.478, p = 0.014) but showed no statistically significant trend with Quality or Flaking Inefficiency.

Surprisingly, grip strength was not predictive of any knapper outcomes in adult males and female participants from both developmental conditions. In child males, grip strength was negatively predictive of Flaking Inefficiency (r(9) = -0.761, p = 0.007) but had no correlation with Quantity or Quality.

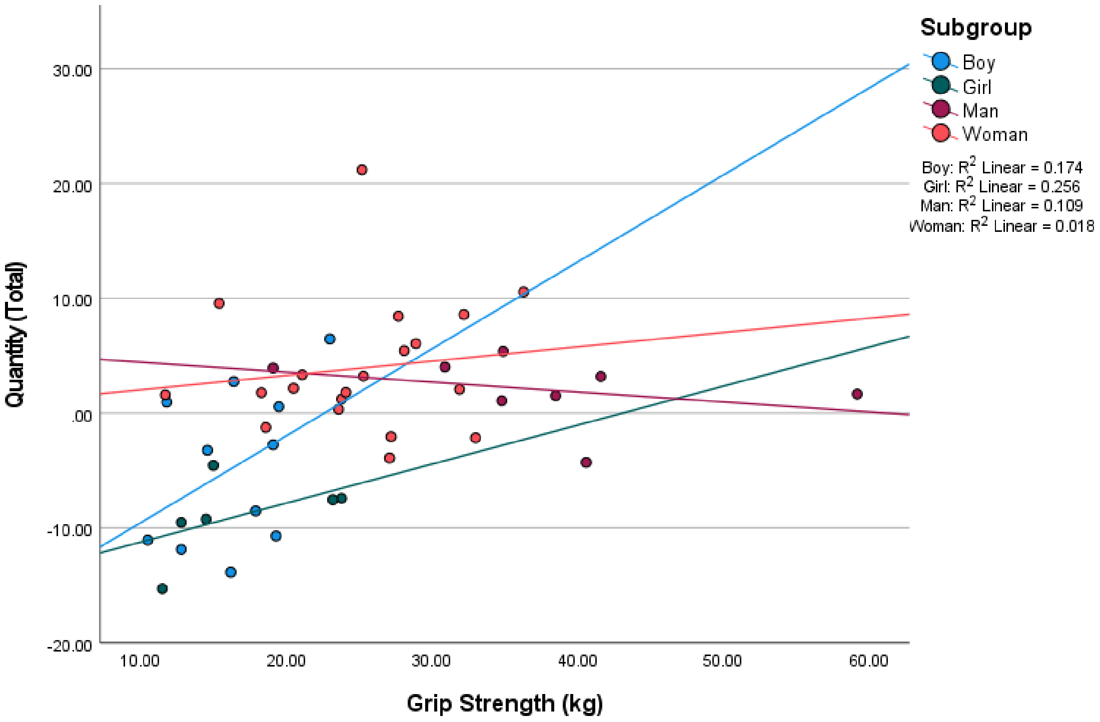


Figure 2.25: Effect of Grip Strength on Lithic Quantity. When all participants were tested together, grip strength positively correlated with lithic Quantity, a trend that was maintained in female participants and marginally in male participants. This trend disappeared within developmental conditions and from individual participant subgroups.

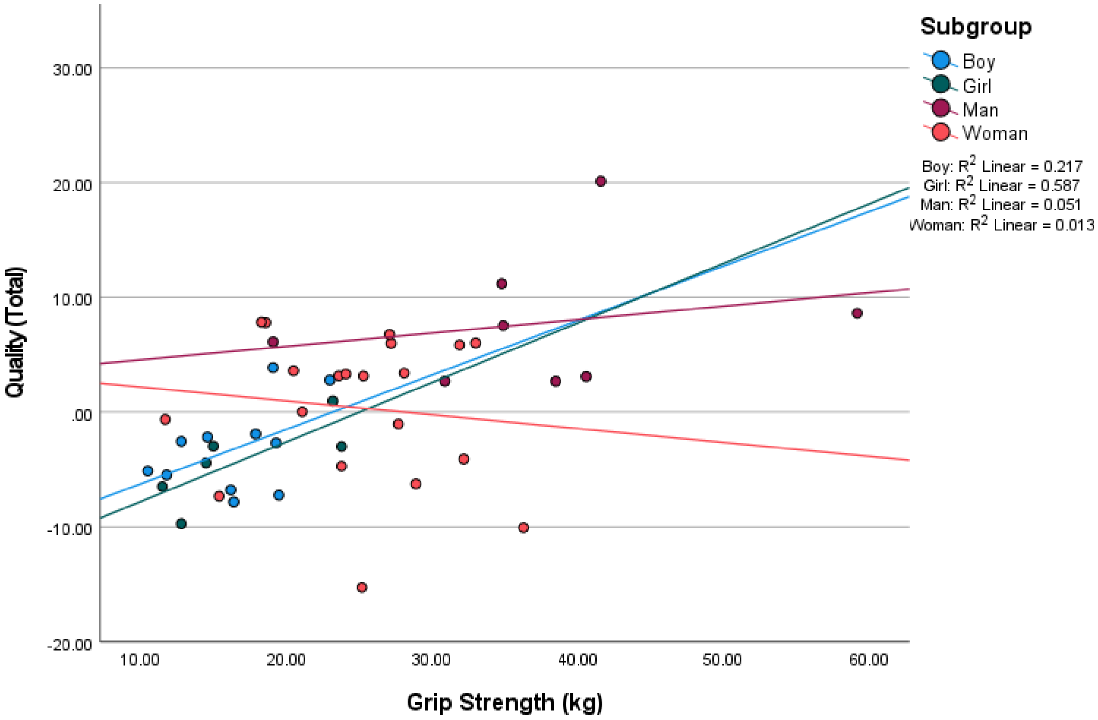


Figure 2.26: Effect of Grip Strength on Lithic Quality. When all participants were tested together, grip strength positively correlated with lithic Quantity, a trend that was maintained in male participants and in children. This trend disappeared within female participants, adults, and from individual participant subgroups.

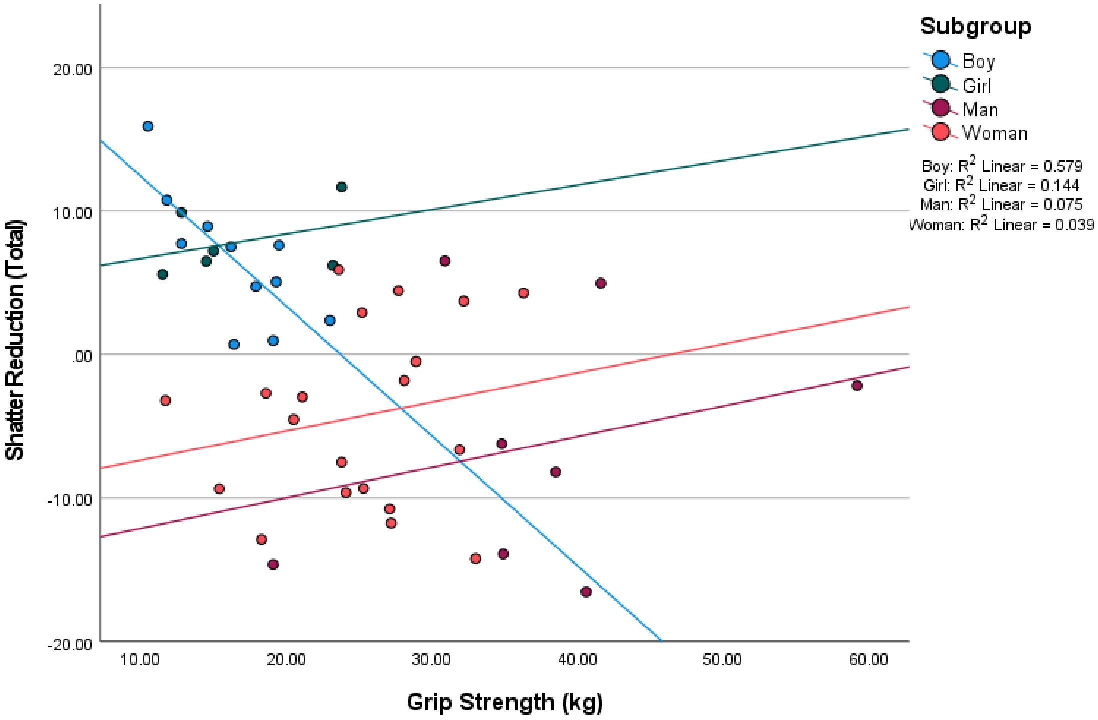


Figure 2.27: Effect of Grip Strength on Flaking Inefficiency. When all participants were tested together, grip strength negatively correlated with lithic Quantity, a trend that was maintained in male participants, generally, and in male children, specifically.

Mental Rotation Task (MRT)

Group-Based Differences

Adult participants' scores on the Mental Rotation Task (MRT) were generally much higher than those of children, at statistical significance ($p < 0.001$) (*Figure 2.30*), and as with grip strength, MRT score increased among child participants with age, at marginal significance ($p = 0.061$) (*Figure 2.29*). Across conditions, there were no differences of MRT performance based on participant sex, nor were there sex-based differences between adult male and female participants and between child male and female participants. These trends were maintained within subgroups (*Figure 2.28*), such that adult male and female participants each outperformed child participants of both sexes.

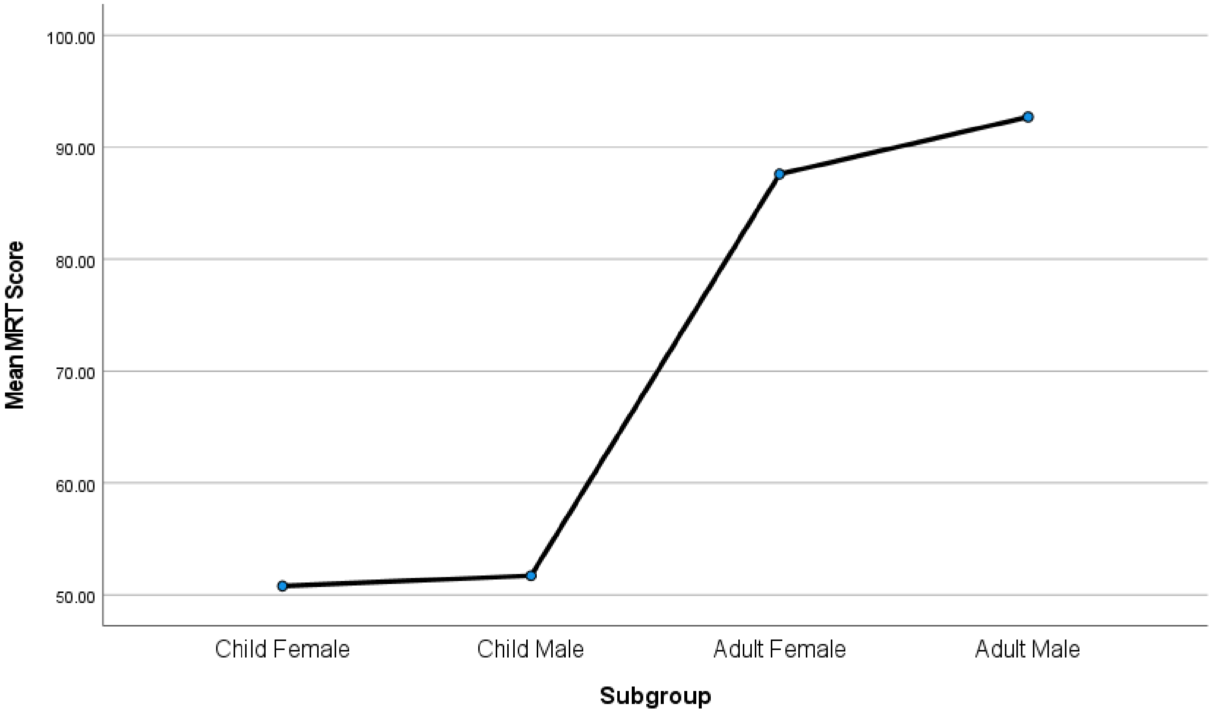


Figure 2.28: Mean MRT Score Across Experimental Subgroups. Adults of both sexes outperformed children of both sexes, but there were no sex-based differences of MRT score within developmental conditions.

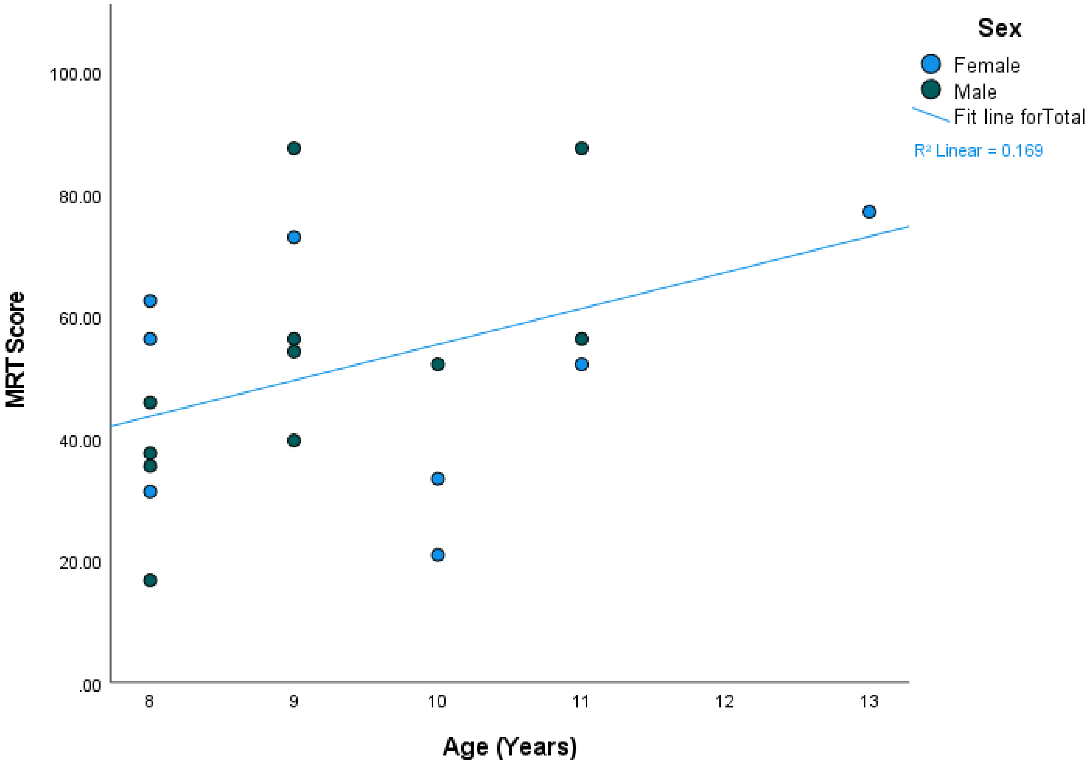


Figure 2.29: Effect of Age on MRT Score in Children. Children's MRT score increased with age, at marginal significance.

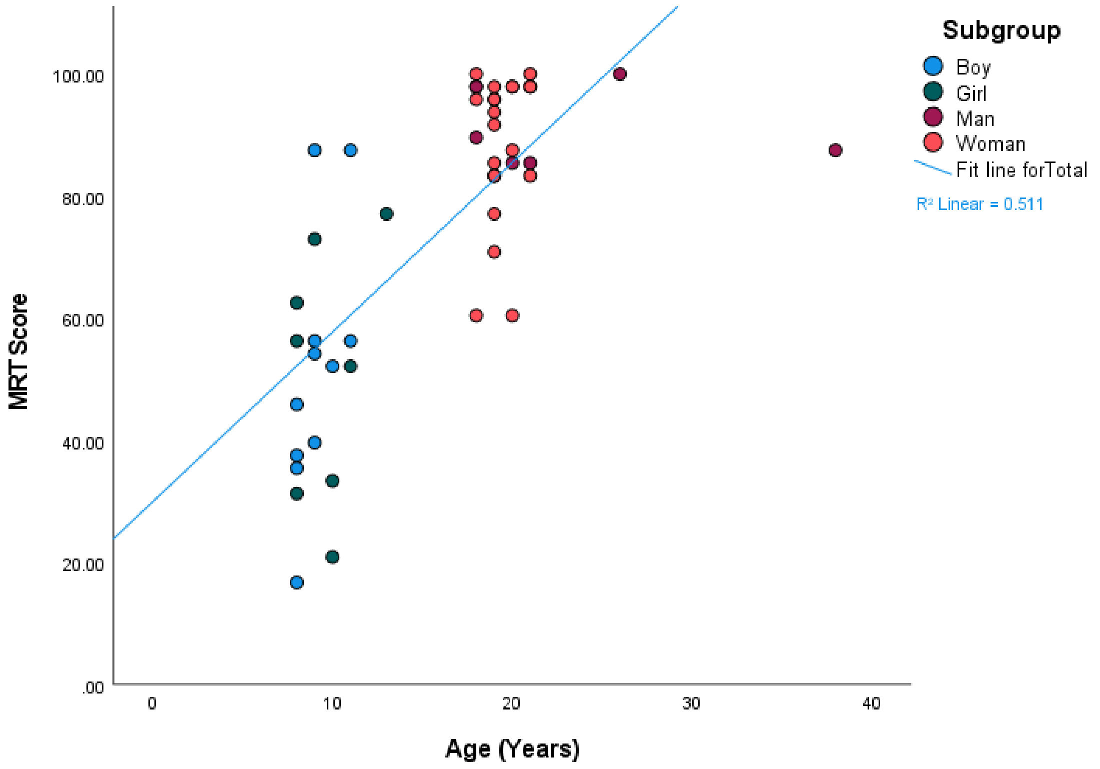


Figure 2.30: Effect of Age on MRT Score in All Participants. MRT score demonstrated a positive relationship with age across developmental conditions.

Correlation with Knapper Skill

When all participants were tested together, performance on the MRT was positively predictive of knapper skill across for lithic Quantity (r(47) = 0.642, p < 0.001) (*Figure 2.31*) and Quality (r(47) = 0.308, p = 0.031) (*Figure 2.32*) and negatively predictive of Flaking Inefficiency (r(47) = -0.592, p < 0.001) (*Figure 2.33*). When children and adults were tested separately, the relationship between MRT score and knapper skill did not yield significant results. However, when male participants were tested independently, MRT scores significantly correlated with all measures of knapper skill, positively, for Quantity (r(17) = 0.722, p < 0.001) and Quality (r(17) = 0.727, p < 0.001), and negatively, for Flaking Inefficiency (r(17) = -0.626, p = 0.004). For female participants, these trends were maintained for Quantity (r(27) = 0.570, $p < 0.001$) and Flaking Inefficiency (r(27) = -0.543, p = 0.002), but there was no statistical correlation for Quality.

All statistical significance fell away when each subgroup was tested, with the exception of a positive correlation of Quantity and MRT score in child males (r(9) = 0.602, p = 0.050).

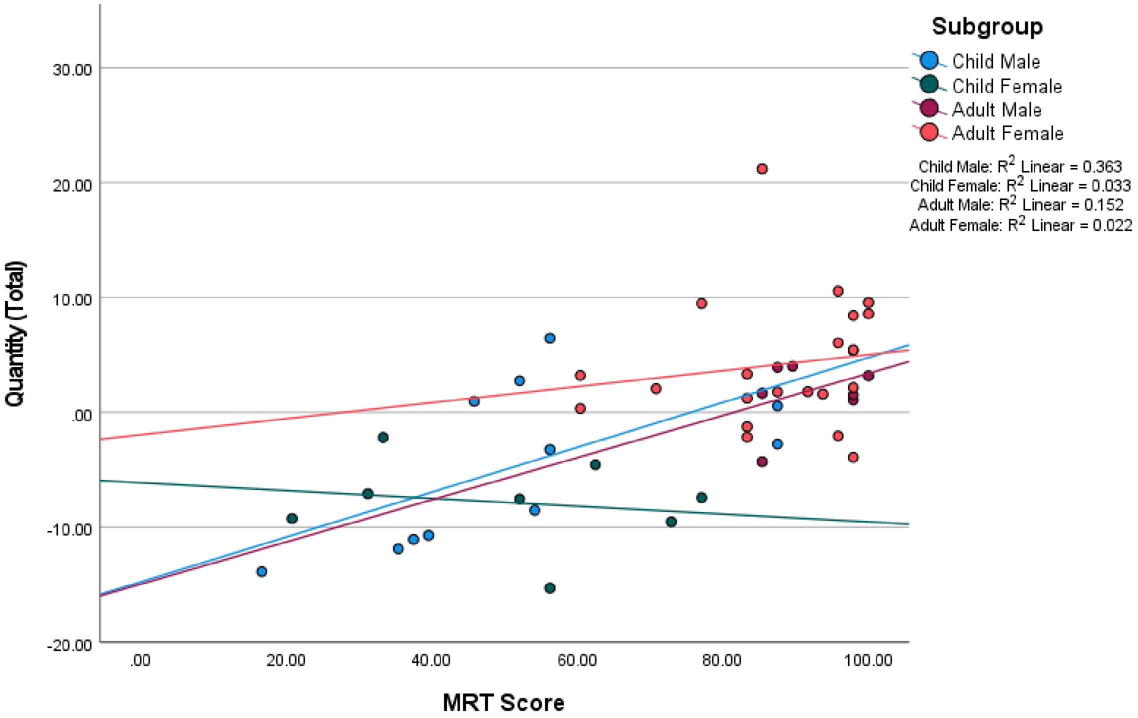


Figure 2.31: Effect of MRT Score on Lithic Quantity. MRT score and lithic Quantity demonstrate a positive relationship, such that with higher MRT scores, participants are more likely to generate a greater quantity of lithic products.

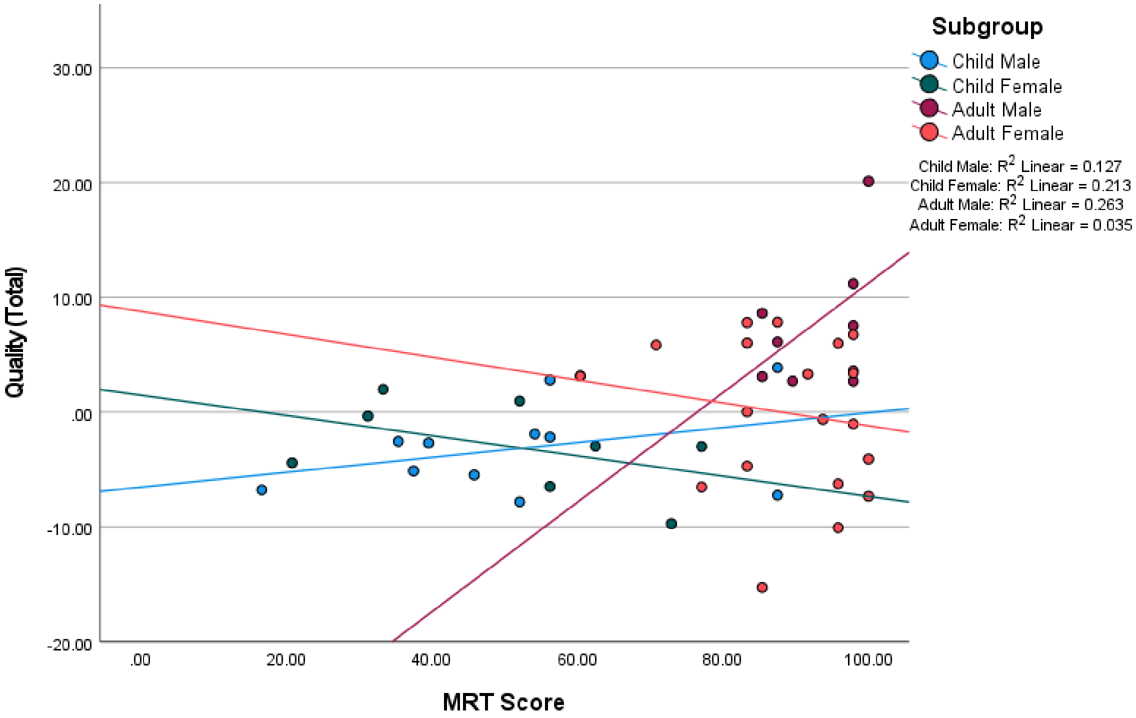


Figure 2.32: Effect of MRT Score on Lithic Quality. Although a somewhat weaker relationship than with lithic Quantity, MRT score and lithic Quality also demonstrated a positive relationship.

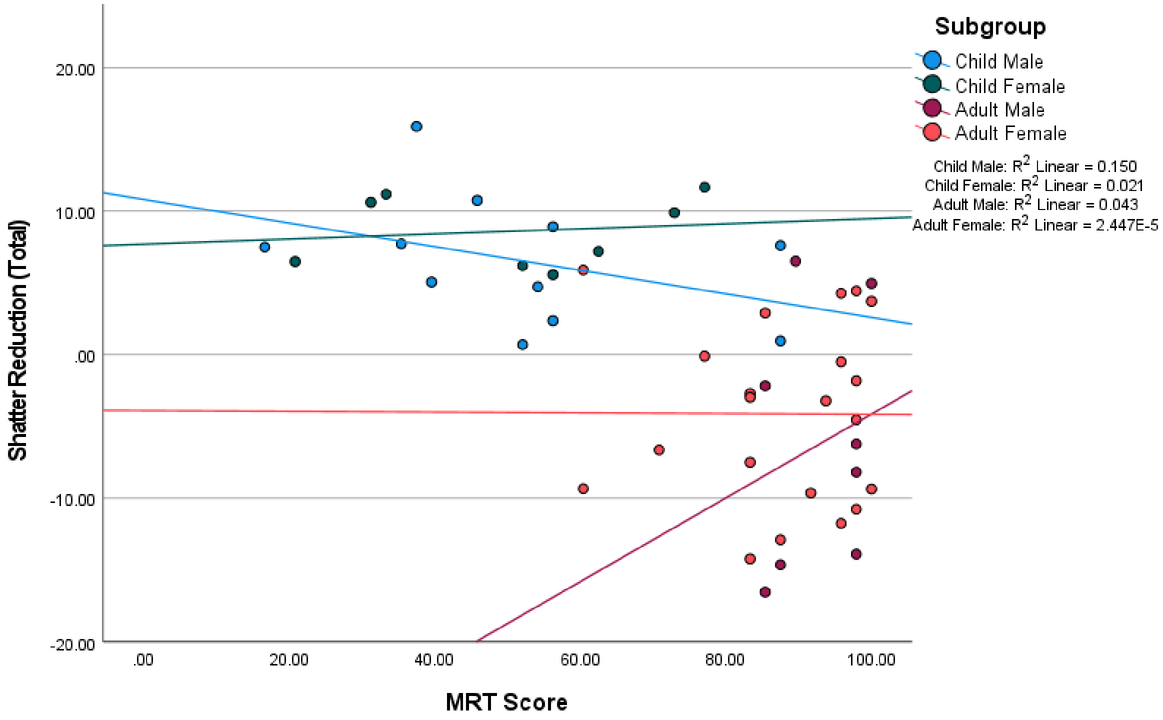


Figure 2.33: Effect of MRT Score on Flaking Inefficiency. MRT score negatively correlated with Flaking Inefficiency, such that participants with higher MRT scores demonstrated more complete core reduction with the production of less shatter.

Fitts' Law Test (FLT)

Group-Based Differences

Adult participants were more likely to complete the Fitts' Law Test (FLT) more quickly ($p < 0.001$) (*Figure 2.34*) and with greater accuracy ($p < 0.001$) (*Figure 2.35*) than child participants. Across conditions, male participants had marginally faster response times than female participants ($p = 0.052$). Within developmental conditions, there were no sex-based differences in response time or accuracy. Although FLT response times and accuracy scores did not vary statistically within children based on age, participant age did demonstrate a positive relationship with FLT performance across both children and adults (*Figures 2.36* and *2.37*).

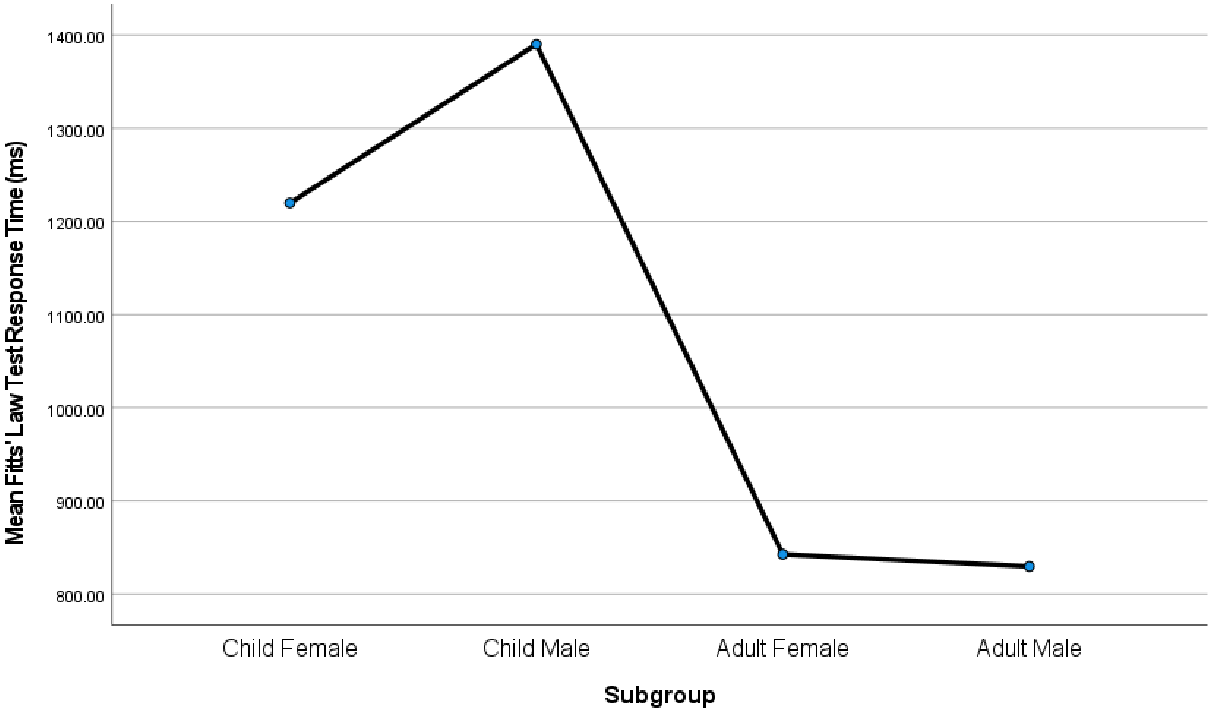


Figure 2.34: Mean Fitts' Law Response Time Across Experimental Subgroups. Adults of both sexes had faster FLT response times than children of both sexes. There was no sex-based difference in performance within developmental conditions.

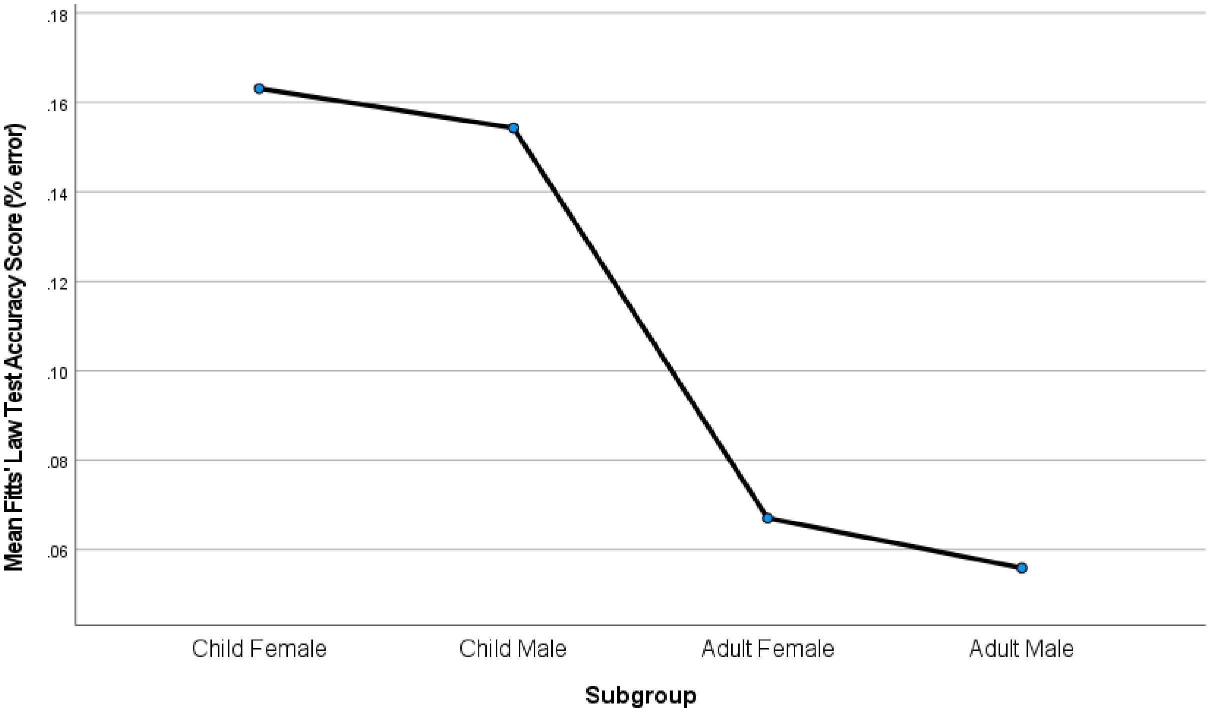


Figure 2.35: Mean Fitts' Law Accuracy Score Across Experimental Subgroups. Adults of both sexes completed the FLT with greater accuracy than children of both sexes. There was no sex-based difference in performance within developmental conditions.

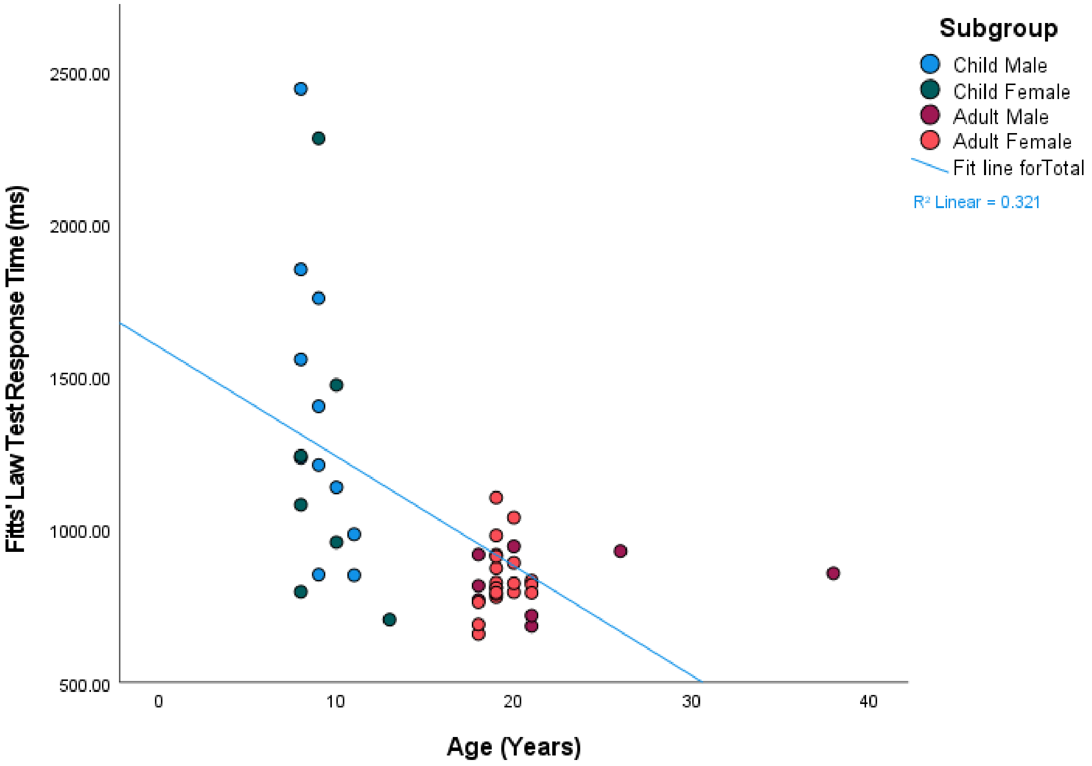


Figure 2.36: Effect of Age on Fitts' Law Response Time in All Participants. Participant FLT response times decreased with age, indicating a better performance.

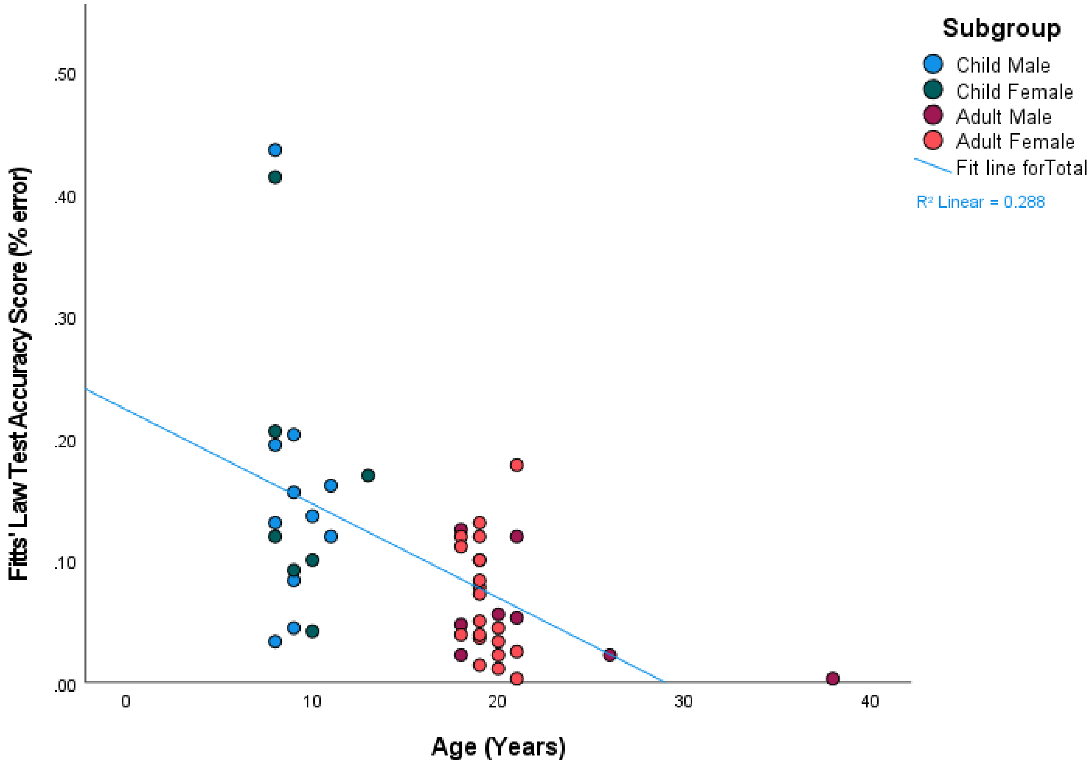


Figure 2.37: Effect of Age on Fitts' Law Accuracy Score in All Participants. The number of errors participants made during the FLT decreased with age.

Correlation with Knapper Skill

With all participants tested together, FLT response times were negatively predictive of lithic Quantity ($r(46) = -0.594$, $p < 0.001$) (*Figure 2.38*) and Quality ($r(46) = -0.352$, $p = 0.014$) (*Figure 2.39*) and positively predictive of Flaking Inefficiency ($r(46) = 0.476$, $p < 0.001$) (*Figure 2.40*). These trends persisted with participants' FLT accuracy scores for Quantity ($r(46) = -0.506$, $p < 0.001$) (*Figure 2.41*) and Flaking Inefficiency ($r(46) = 0.397$, $p = 0.005$) (*Figure 2.42*), but there was no statistically significant correlation for Quality ($p = 0.087$). Although knapper skill metrics were not correlated with Fitts' accuracy scores or response time in adult participants, response time was negatively predictive of lithic Quantity in children ($r(16) = -0.495$, $p = 0.037$). As with the MRT scores, significant correlations appeared when male participants were tested separately, both for response time and accuracy. Specifically, response time was negatively correlated with Quantity ($r(17) = -0.831$, $p < 0.001$) and Quality ($r(17) = -0.578$, $p = 0.010$) and positively correlated with Flaking Inefficiency ($r(17) = 0.542$, $p = 0.016$). These trends persisted with FLT accuracy for Quality ($r(17) = -0.534$, $p = 0.019$) and Flaking Inefficiency ($r(17) = 0.492$, $p = 0.032$), but there was no significant correlation with Quantity in male participants. In female participants, FLT response time negatively correlated with lithic Quantity ($r(26) = 0.382$, $p = 0.045$) and positively with knapper Trade-Off ($r(26) = 0.377$, $p = 0.048$) but had no statistical relationship to lithic Quality. FLT accuracy was negatively associated with Quantity ($r(26) = -0.612$, $p < 0.001$) but had no correlation with Quantity or Flaking Inefficiency.

Also similar to the MRT results, all significant correlations disappeared when each subgroup was tested separately, with the exception of male children, whose FLT response times negatively corresponded with lithic Quantity ($r(9) = -0.795$, $p = 0.003$).

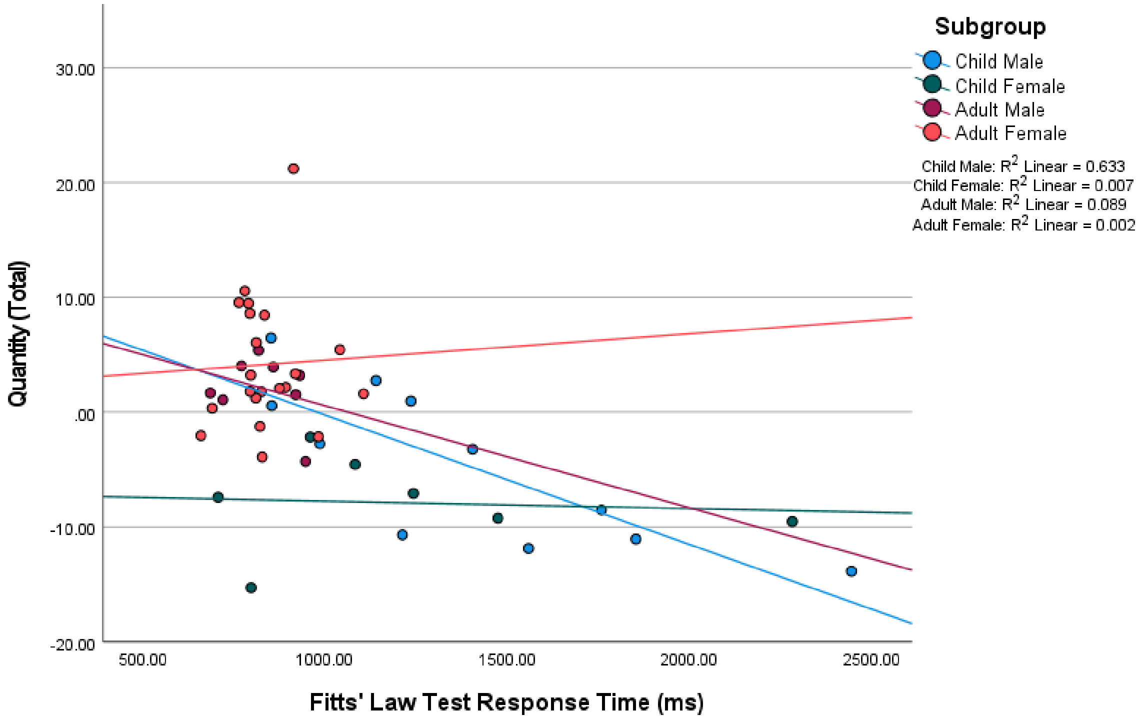


Figure 2.38: Effect of Fitts' Law Response Time on Lithic Quantity. Participants with quicker FLT response times were more likely to produce a greater Quantity of lithics.

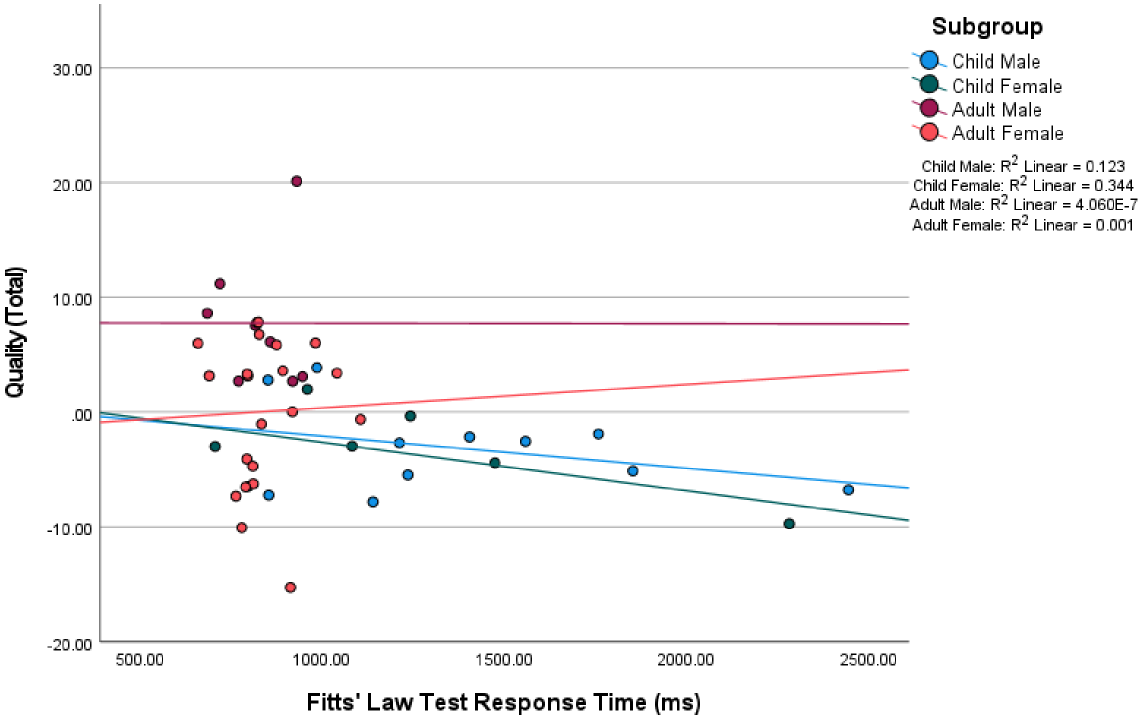


Figure 2.39: Effect of Fitts' Law Response Time on Lithic Quality. Participants with quicker FLT response times were more likely to produce lithics of a greater Quality.

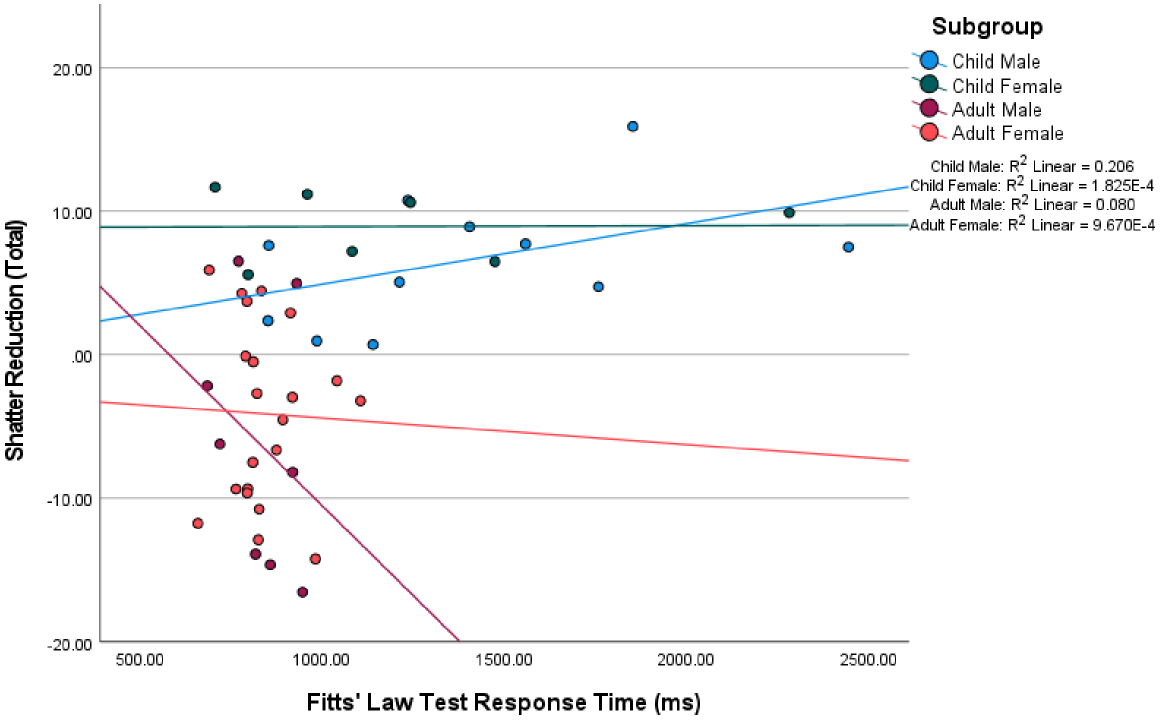


Figure 2.40: Effect of Fitts' Law Response Time on Flaking Inefficiency. Participants with quicker FLT response times were more likely to more completely reduce their cores and less likely to produce shatter.

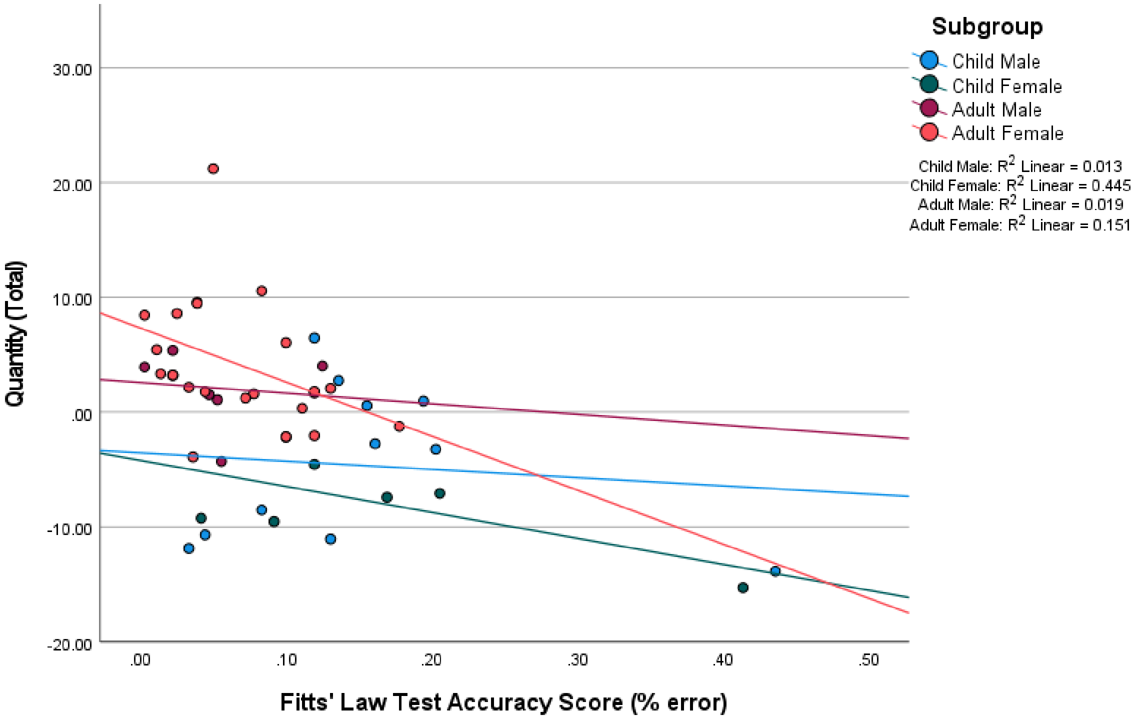


Figure 2.41: Effect of Fitts' Law Accuracy Score on Lithic Quantity. Participants with fewer FLT errors were more likely to produce a greater Quantity of lithics.

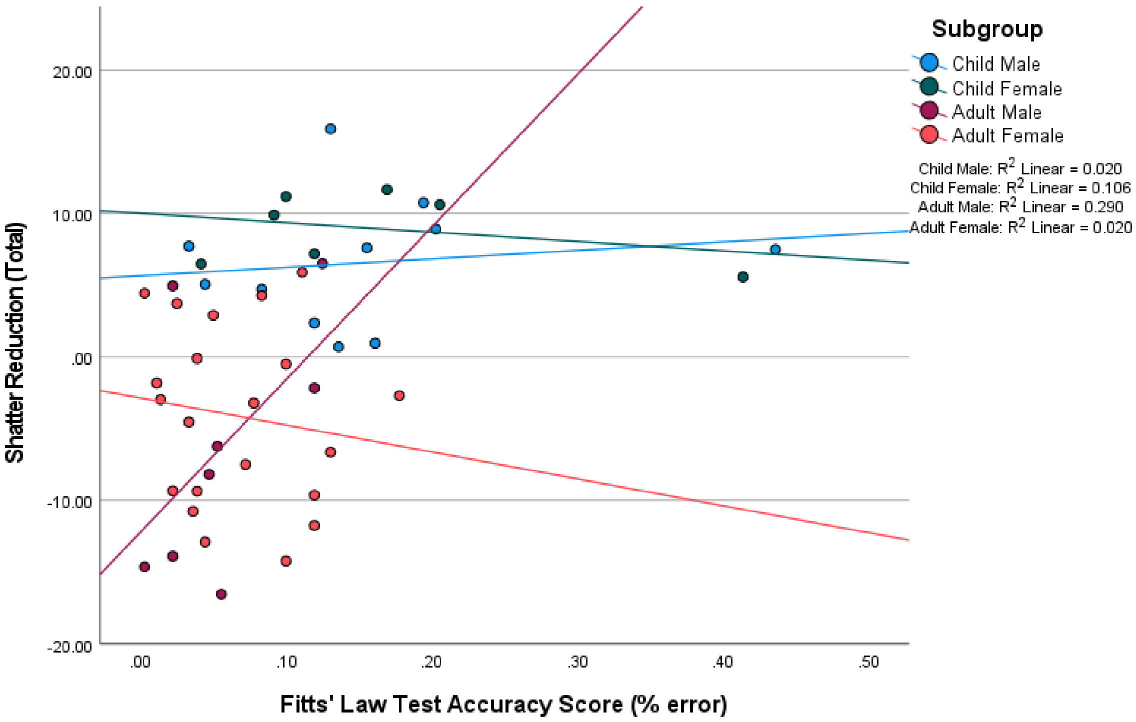


Figure 2.42: Effect of Fitts' Law Response Time on Flaking Inefficiency. Participants with fewer FLT errors were more likely to more completely reduce their cores and to produce less shatter.

Stroop Raw Color-Word Test

Group-Based Differences

Adults and children demonstrated no statistically significant difference in their Stroop Raw Color-Word Test scores, nor did male and female participants. Within conditions, adult male and female participants also showed no difference in their Stroop scores, and among children, girls marginally outperformed boys (p = 0.065).

Correlation with Knapper Skill

When all participants were tested together, Stroop scores were not predictive of any knapping skill metrics. Children and male participants maintained this trend when tested separately, and adults showed a marginal positive correlation with lithic Quantity (r(28) = 0.347, p = 0.06). Female participants demonstrated a positive relationship between Stroop scores and core Flaking Inefficiency (r(27) = 0.442, p = 0.019) (*Figure 2.43*). When examined separately, none of the four experimental subgroups revealed significant relationships between Stroop scores and knapper skill.

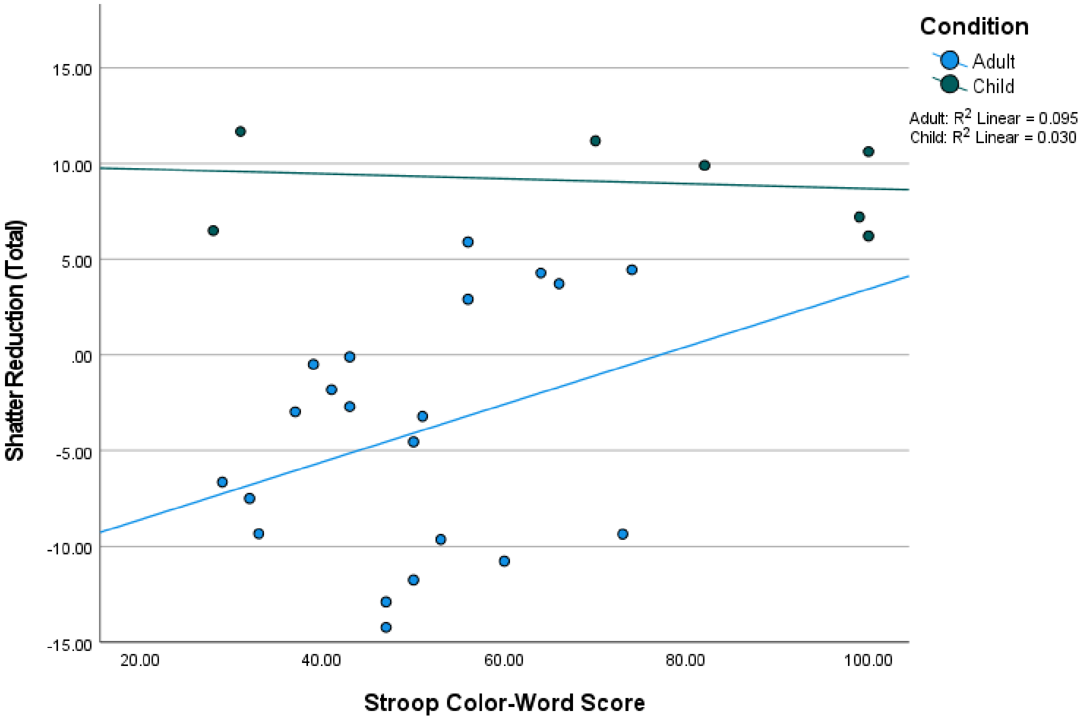


Figure 2.43: Effect of Stroop Color-Word Score on Flaking Inefficiency in Female Participants. Female participants with higher Stroop scores were more likely to reduce their cores less completely and to produce a greater amount of shatter.

Big Five Inventory (BFI)

Group-Based Differences

The Big Five Inventory (BFI) is designed to evaluate individuals on five personality characteristics that are demonstrated to remain stable across a person's lifetime: Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism. Adult and child participants displayed no differences in their scoring for any of the BFI's measured traits; however, male and female participants showed differences in Openness ($p = 0.02$) (*Figure 2.44*) and Neuroticism ($p = 0.046$) (*Figure 2.45*), with males, on average, having higher Openness scores and females having higher scores in Neuroticism. Within conditions, adults showed no sex-based differences, and among children, the trend for Openness in males persisted, with boys scoring higher than girls ($p = 0.002$).

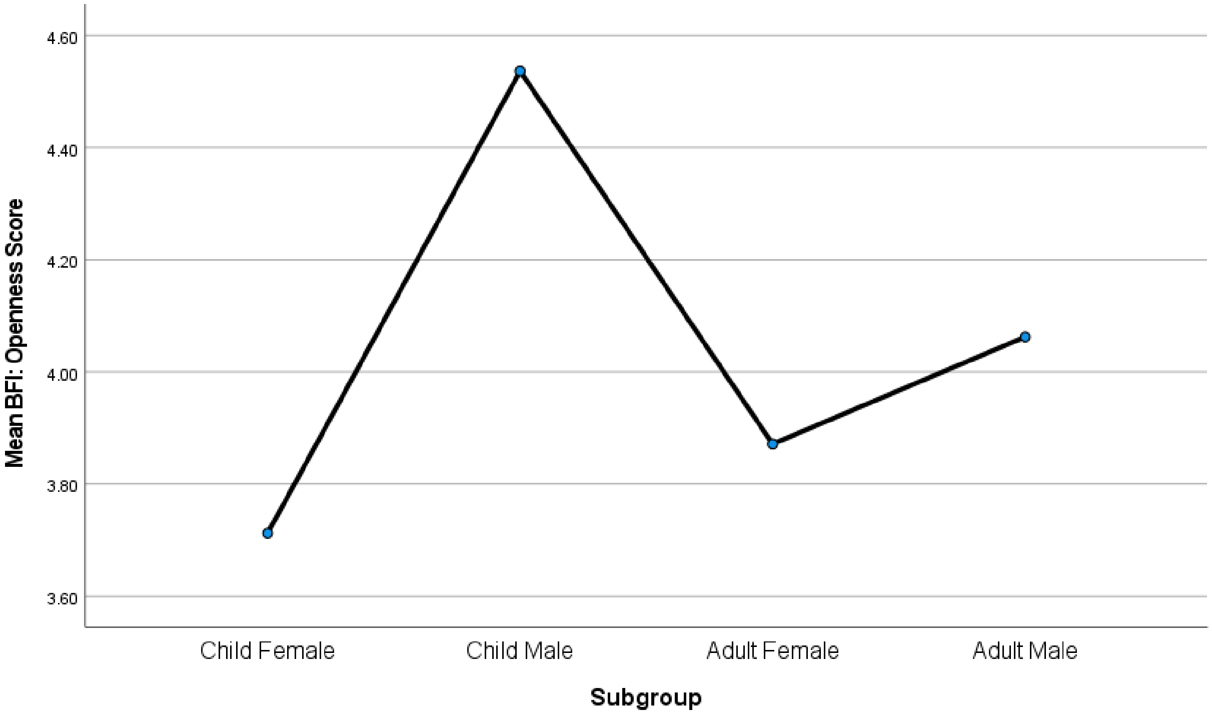


Figure 2.44: Mean BFI: Openness Score Across Experimental Subgroups. On average, male participants reported higher Openness scores than female participants.

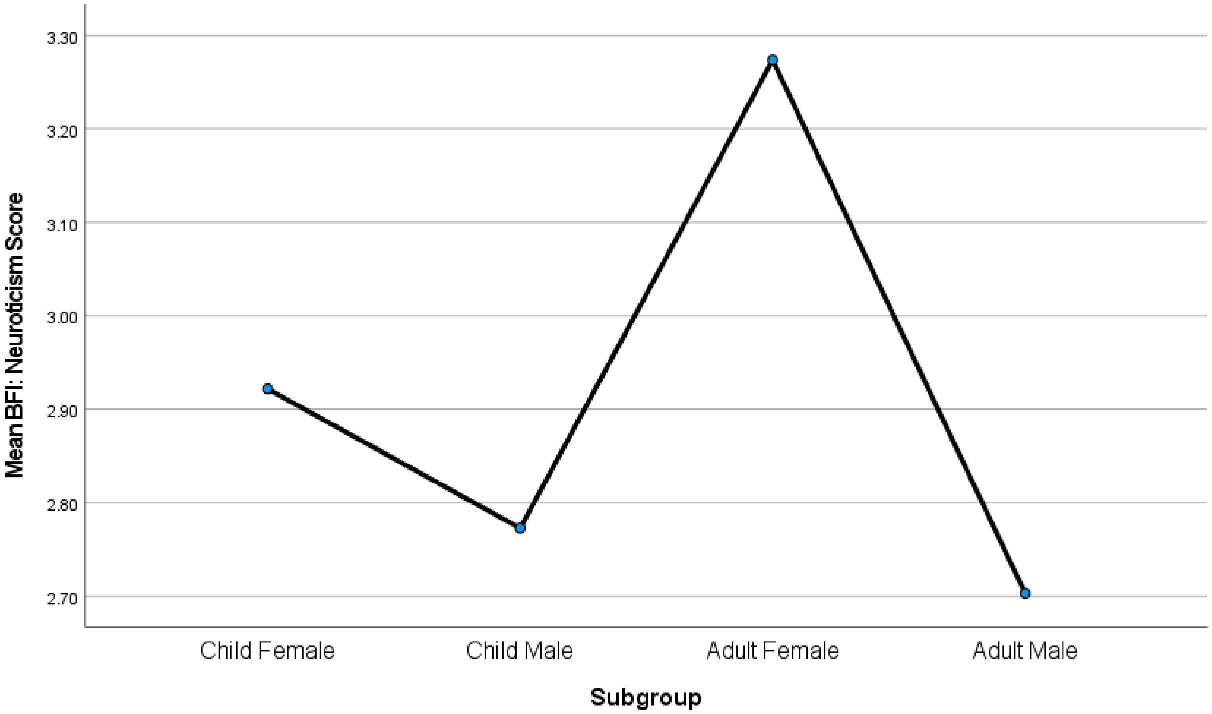


Figure 2.45: Mean BFI: Neuroticism Score Across Experimental Subgroups. On average, female participants reported higher Neuroticism scores than male participants.

Correlation with Knapper Skill

When all participants were tested together, none of the BFI traits were predictive of knapper outcomes. In children, higher Agreeableness scores negatively corresponded with greater lithic Quantity (r(17) = -0.539, p = 0.017) (*Figure 2.46*). In adults, higher scores in Agreeableness were negatively predictive of Flaking Inefficiency (r(28) = -0.367, p = 0.046) (*Figure 2.47*). In male participants, Openness scores negatively correlated with Quantity (r(17) = -0.540, p = 0.017) (*Figure 2.48*) and positively with Flaking Inefficiency (r(17) = 0.571, p = 0.011) (*Figure 2.49*). Among female participants, BFI scores were not predictive of Quantity, Quality, or Flaking Inefficiency.

For adult males, Conscientiousness positively correlated with lithic Quantity (r(6) = 0.881, p = 0.004) (*Figure 2.50*), and Agreeableness negatively corresponded to Flaking Inefficiency (r(6) = -0.867, p = 0.005) (*Figure 2.51*). None of the BFI traits were predictive of knapper skill in adult females, and in child females, only Neuroticism was positively predictive of lithic Quantity (r(6) = 0.773, p = 0.039) (*Figure 2.52*). In child males, Openness (r(9) = -0.690, p = 0.019) (*Figure 2.53*), Conscientiousness (r(9) = -0.635, p = 0.036) (*Figure 2.54*), and Agreeableness (r(11) = -0.705, p = 0.015) (*Figure 2.55*) were all negatively correlated with Quantity.

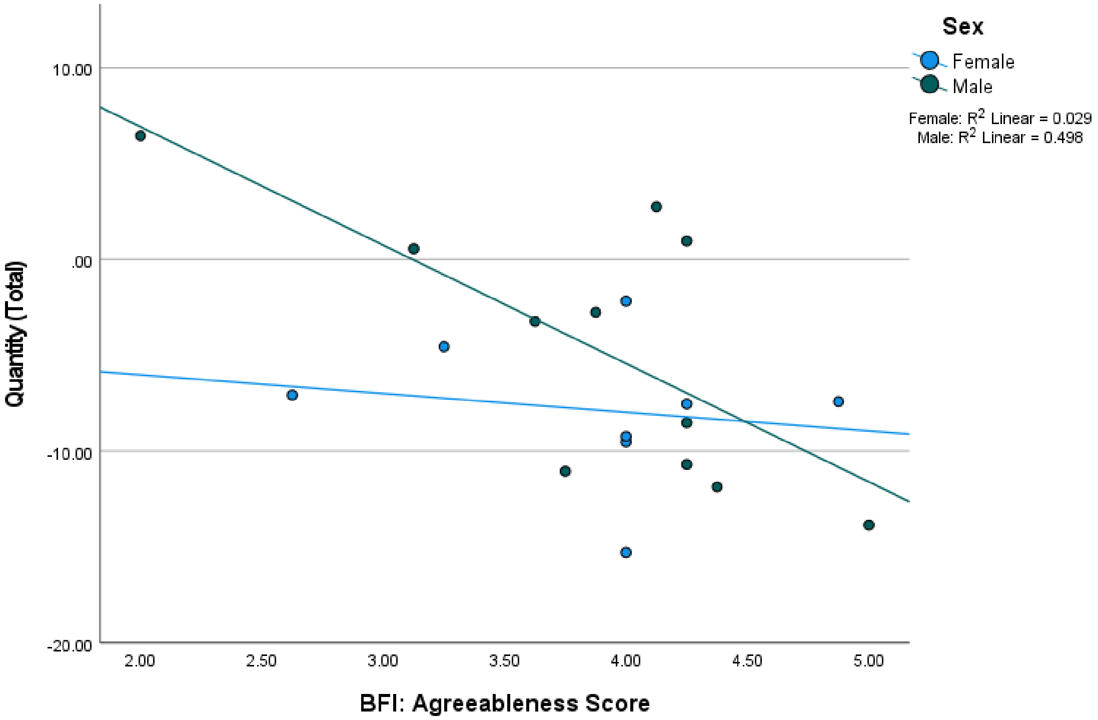


Figure 2.46: Effect of BFI: Agreeableness Score on Lithic Quantity in Children. Child participants who scored higher on the BFI: Agreeableness factor were more likely to produce fewer lithics.

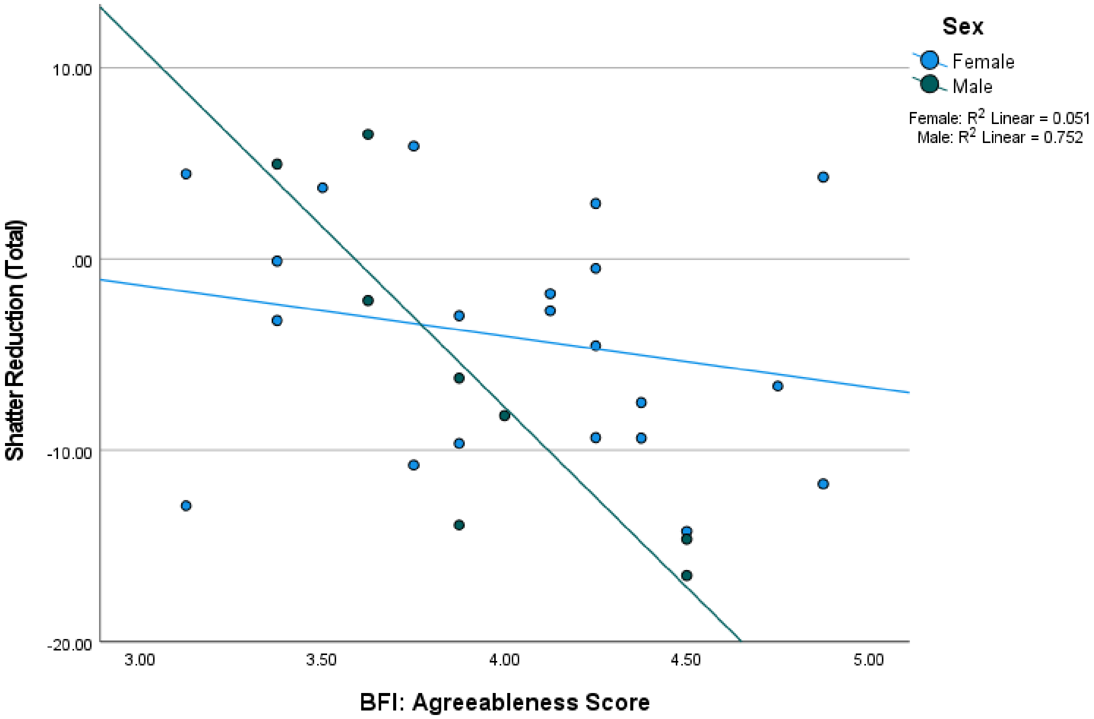


Figure 2.47: Effect of BFI: Agreeableness Score on Flaking Inefficiency in Adults. Adult participants who scored higher on the BFI: Agreeableness factor were more likely to more completely reduce their cores and to produce less shatter.

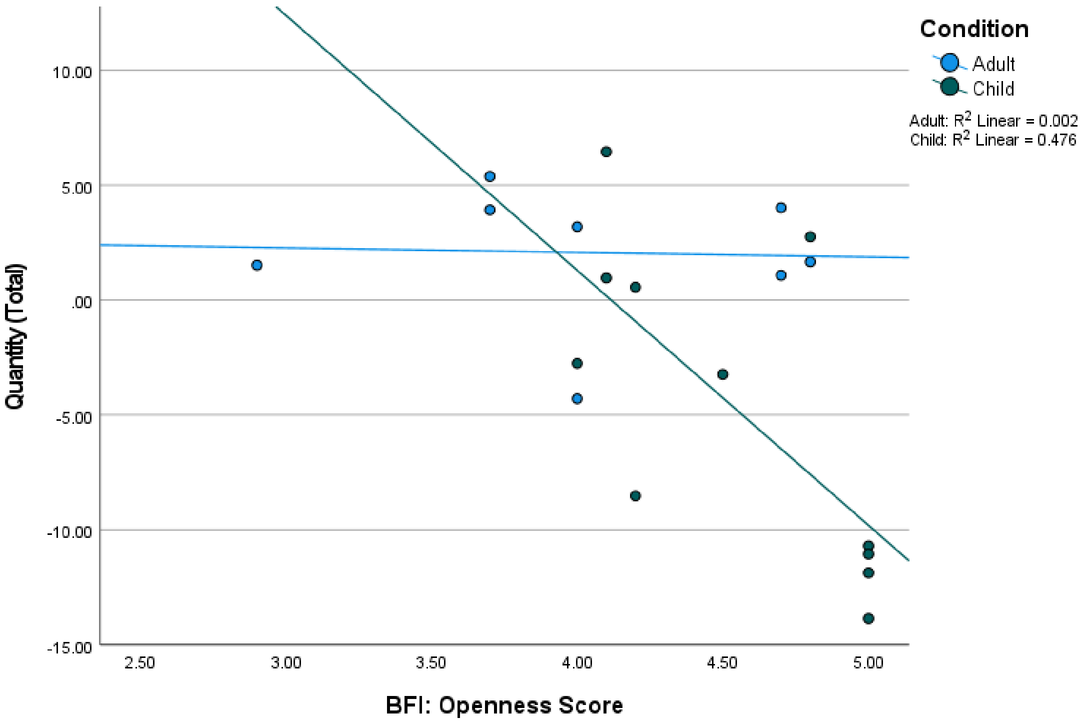


Figure 2.48: Effect of BFI: Openness Score on Lithic Quantity in Male Participants. Male participants who scored higher on the BFI: Openness factor were more likely to produce fewer lithics.

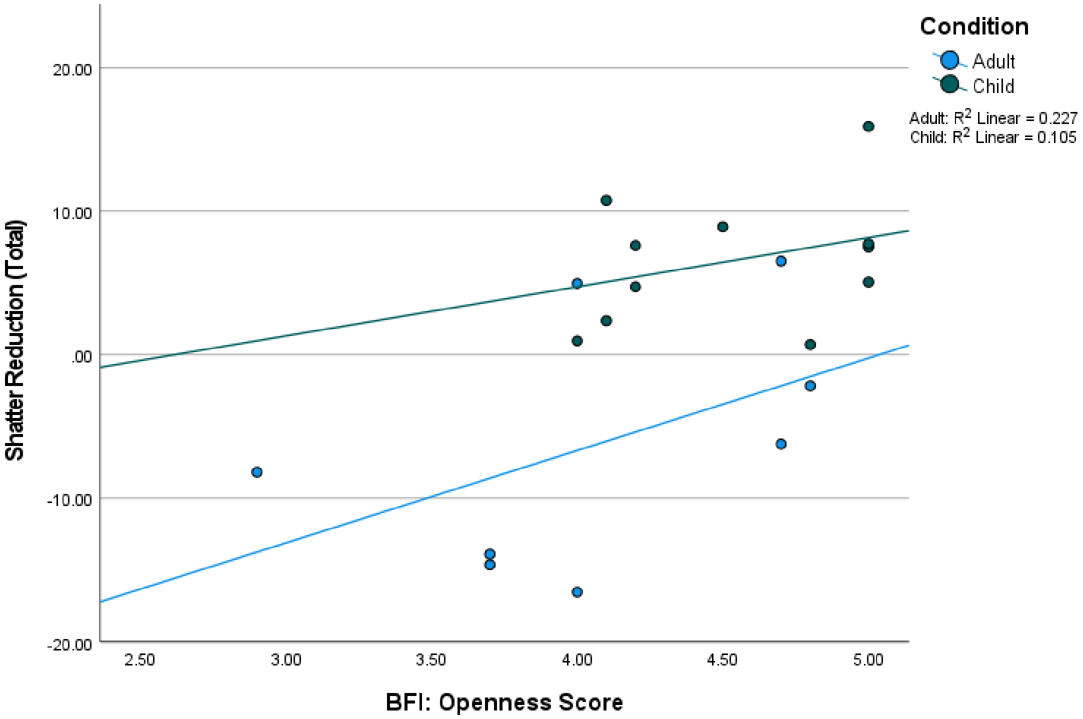


Figure 2.49: Effect of BFI: Openness Score on Flaking Inefficiency in Male Participants. Male participants who scored higher on the BFI: Openness factor were more likely to less intensely reduce their cores and to produce more shatter.

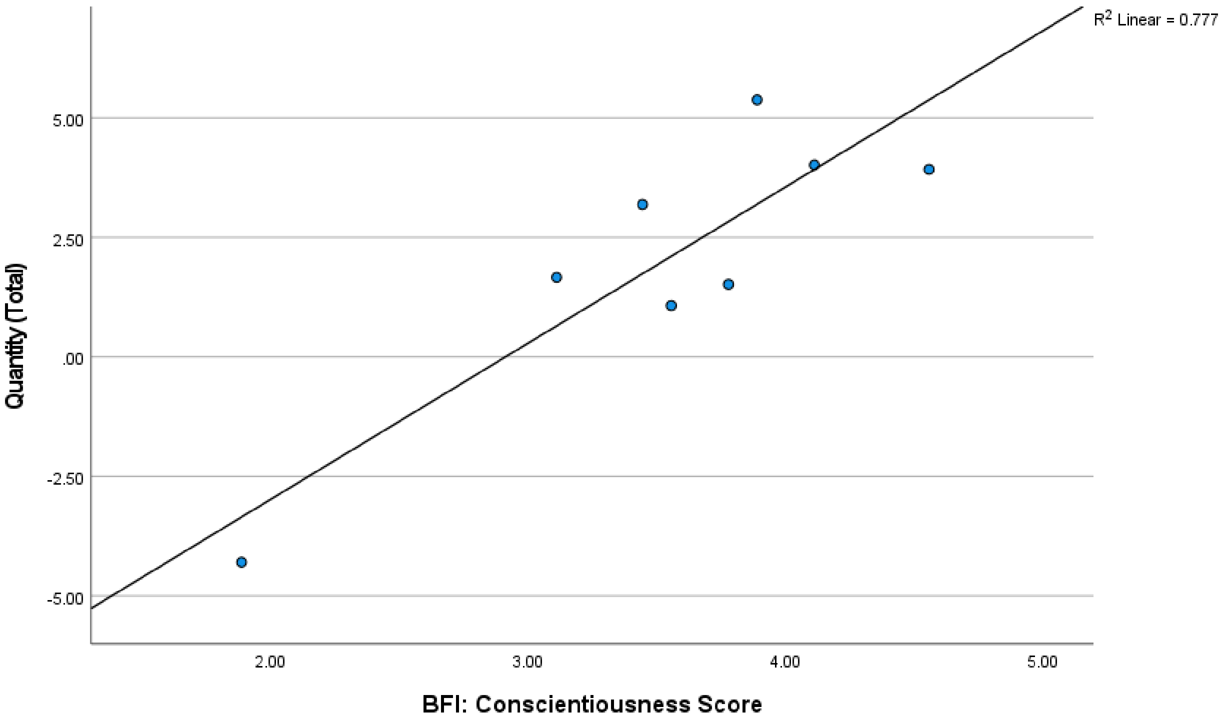


Figure 2.50: Effect of BFI: Conscientiousness Score on Lithic Quantity in Adult Male Participants. More conscientious adult male participants were more likely to produce a greater number of lithics.

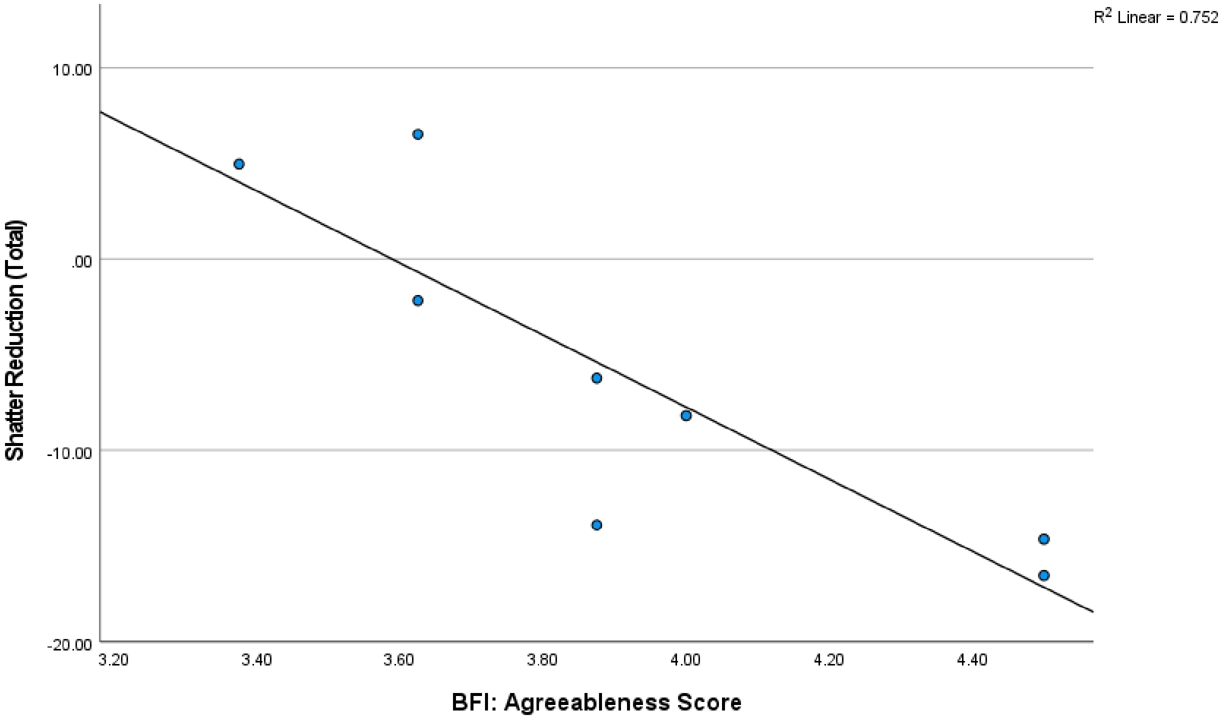


Figure 2.51: Effect of BFI: Agreeableness Score on Flaking Inefficiency in Adult Male Participants. More agreeable adult male participants were more likely to more intensely reduce their cores and to produce less shatter.

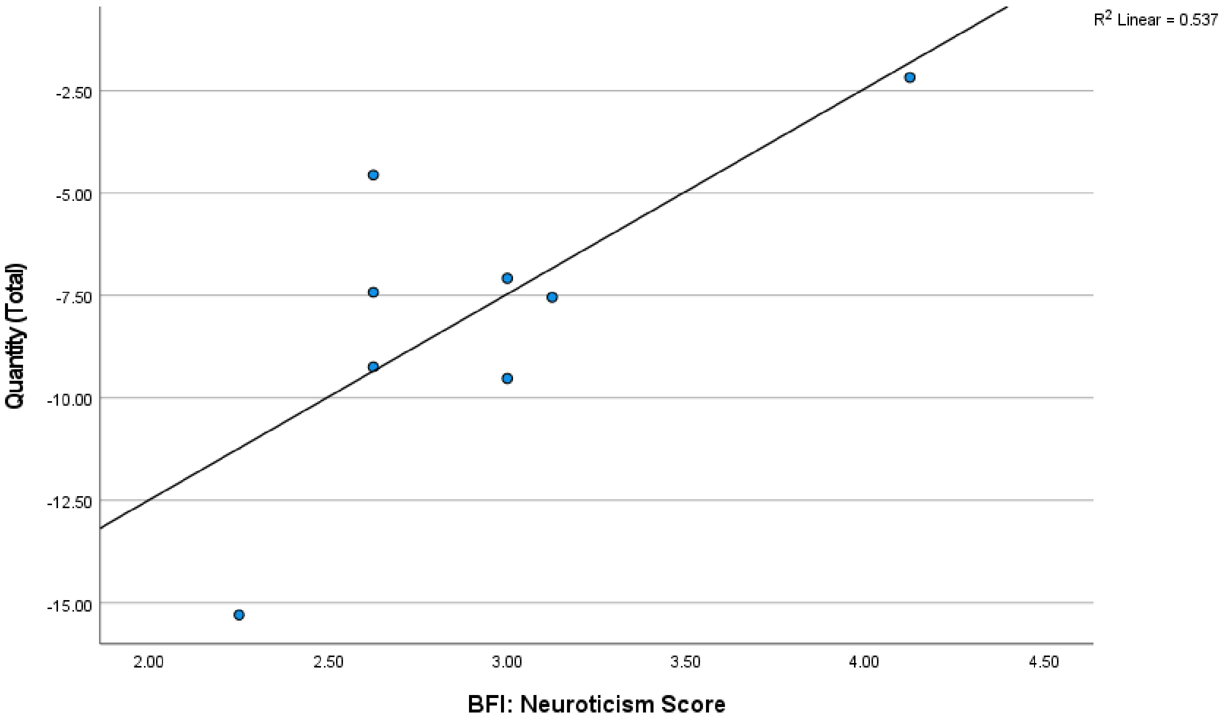


Figure 2.52: Effect of BFI: Neuroticism Score on Lithic Quantity in Child Female Participants. More neurotic child female participants were more likely to produce a greater quantity of lithics.

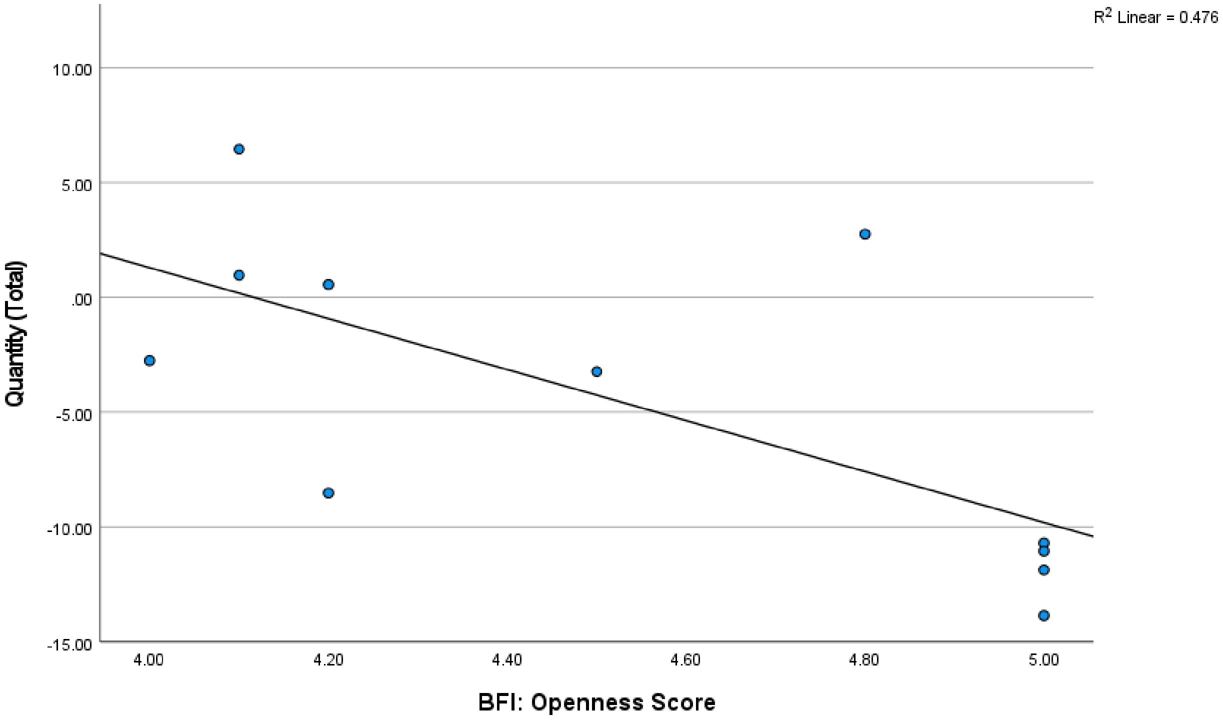


Figure 2.53: Effect of BFI: Openness Score on Lithic Quantity in Child Male Participants. More open child male participants produced fewer lithics.

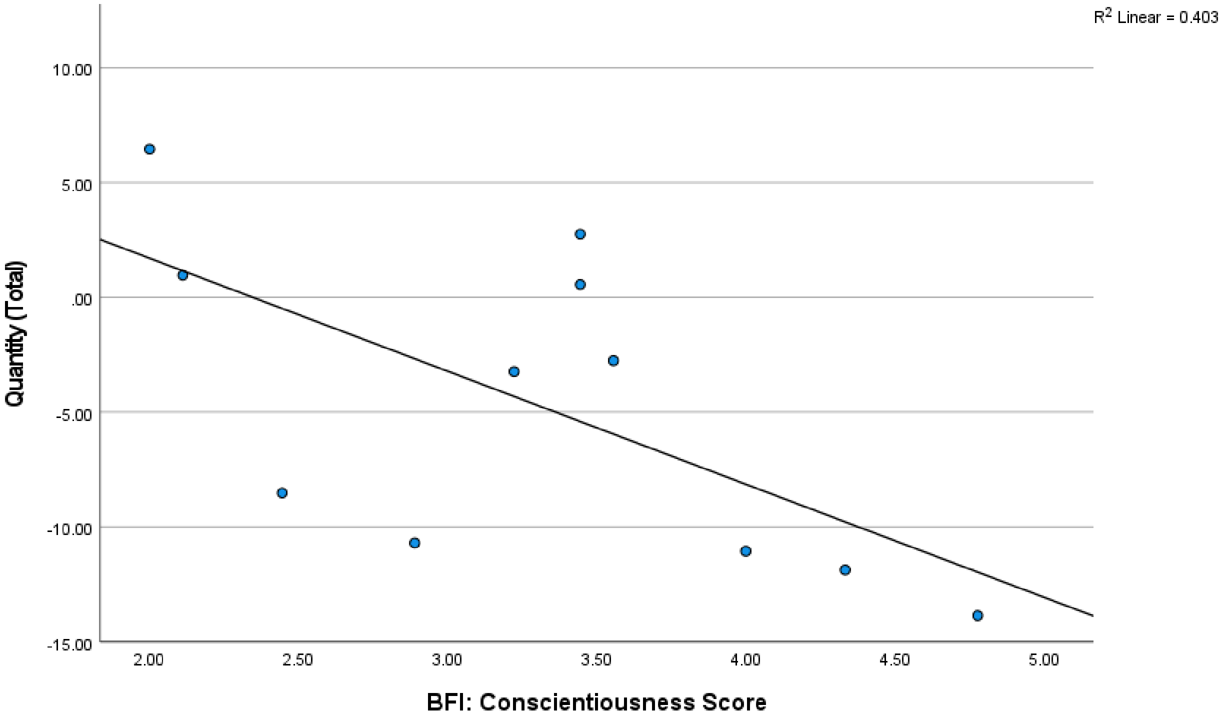


Figure 2.54: Effect of BFI: Conscientiousness Score on Lithic Quantity in Child Male Participants. More conscientious child male participants produced fewer lithics.

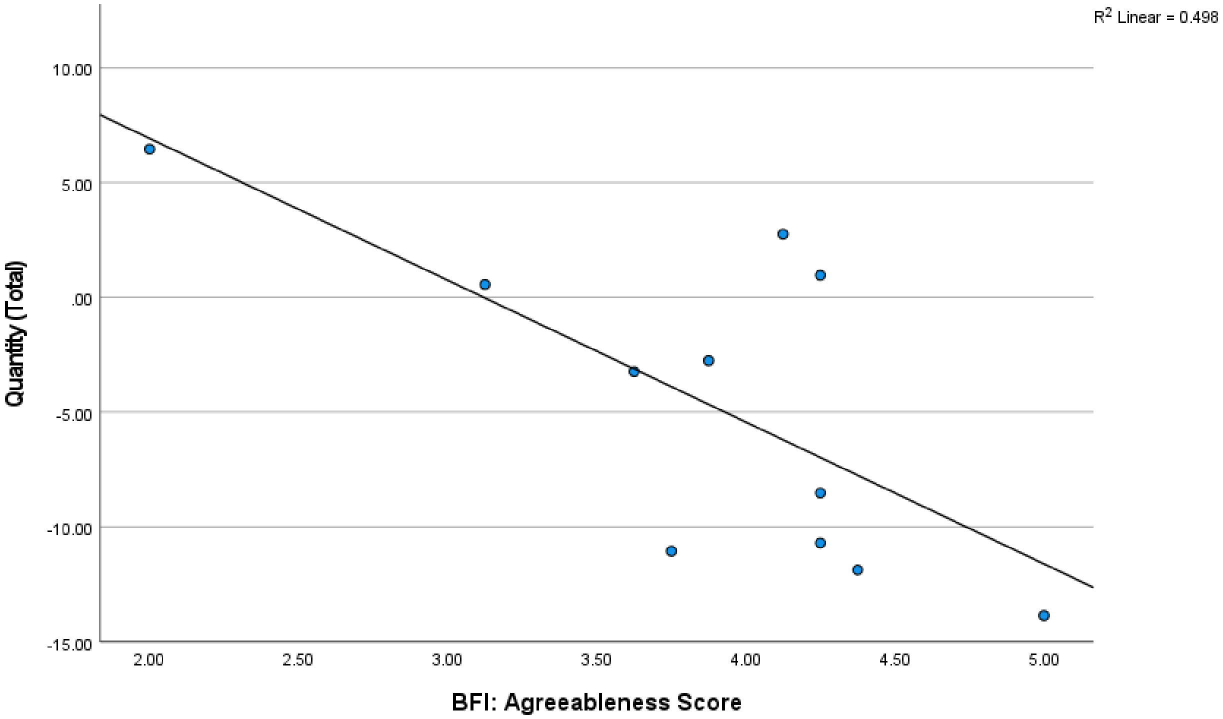
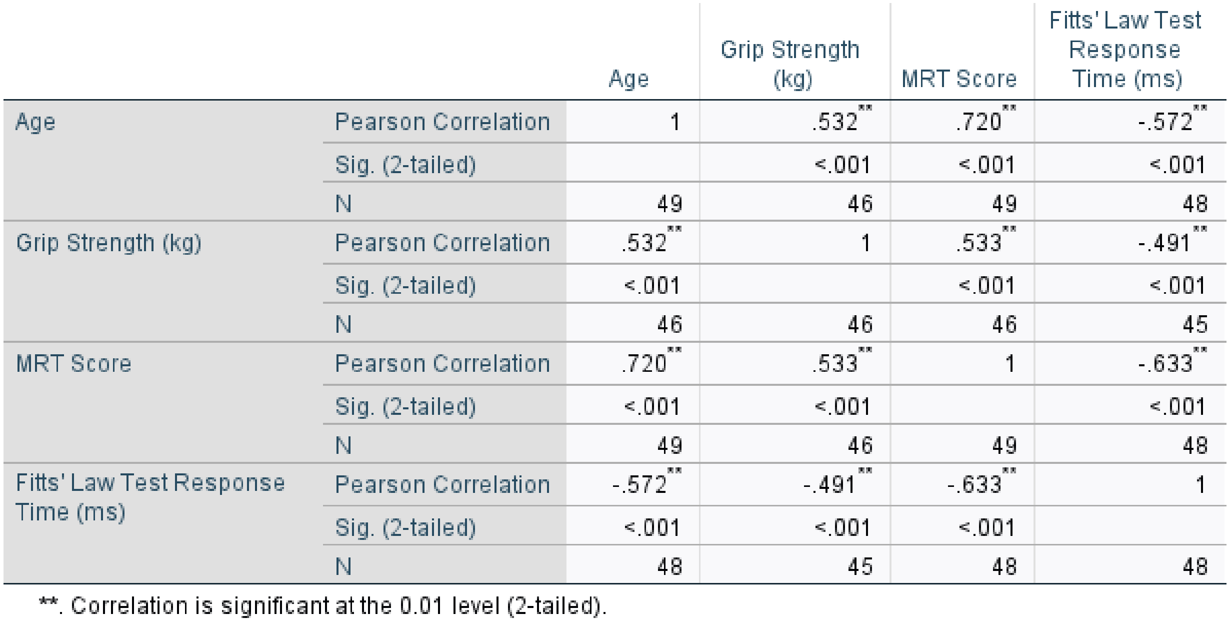


Figure 2.55: Effect of BFI: Agreeableness Score on Lithic Quantity in Child Male Participants. More agreeable child male participants produced fewer lithics.

Stepwise Regressions

The above analyses have highlighted the importance of the following variables to knapper outcomes: participant age, grip strength, MRT score, and FLT response times. To examine whether they act independently of each other to influence knapper skill, I ran stepwise multiple regressions of these variables against Quantity, Quality, and Flaking Inefficiency. The models yielded from this analysis, when all participants were included, all performed at the level of significance (Quantity: $p < 0.001$; Quality: $p = 0.003$; Flaking Inefficiency: $p < 0.001$). However, the psychometric and motor coefficients included in the model largely failed to contribute to their respective models at the level of significance. Namely, for Quantity, participant MRT score came the closest to a significant value at $p = 0.078$. For Quality, the model also was significant at p = 0.003, with grip strength (p = 0.058) and age (p = 0.052) acting as the most salient coefficients. The Flaking Inefficiency model (p < 0.001) was the only model to include a statistically significant coefficient: age, at p = 0.006. As such, participant age, grip strength, MRT score, and FLT response times demonstrated a high degree of multicollinearity, thus making it difficult to identify how each coefficient contributed to the models. This result is a logical one given the highly significant level of correlation between these variables (*Table 2.5*).

Table 2.5: Pearson Correlations Between Age, Grip Strength, MRT Score, and Fitts' Law Test Response Time in All Participants



		Age	Grip Strength (kg)	MRT Score	Fitts' Law Test Response Time (ms)
Age	Pearson Correlation	1	$.532^{**}$	$.720^{**}$	$-.572^{**}$
	Sig. (2-tailed)		<.001	<.001	<.001
	N	49	46	49	48
Grip Strength (kg)	Pearson Correlation	$.532^{**}$	1	$.533^{**}$	$-.491^{**}$
	Sig. (2-tailed)	<.001		<.001	<.001
	N	46	46	46	45
MRT Score	Pearson Correlation	$.720^{**}$	$.533^{**}$	1	$-.633^{**}$
	Sig. (2-tailed)	<.001	<.001		<.001
	N	49	46	49	48
Fitts' Law Test Response Time (ms)	Pearson Correlation	$-.572^{**}$	$-.491^{**}$	$-.633^{**}$	1
	Sig. (2-tailed)	<.001	<.001	<.001	
	N	48	45	48	48

**. Correlation is significant at the 0.01 level (2-tailed).

Tested separately, none of the four participant subgroups yielded significant models, nor did those run with only adult participants or only children. Regarding participant sex, male participants, both adults and children, demonstrated particularly strong relationships between knapper outcomes and the above psychometric and motor variables. When these same stepwise multiple regression analyses were conducted with only male participants, the lithic Quantity model yielded a significant result (p = 0.001), with FLT response time as a significant coefficient (p = 0.011). The Quality model was significant at p = 0.001, and as with the previous model that included all participants, age (p = 0.071) and grip strength (p = 0.094) contributed the most to the model outcomes, though not at the level of significance. Finally, the Flaking Inefficiency modal was significant at p = 0.022, with age contributing to the model at p = 0.079. Stepwise regression models generated for female participants reveled significant values for lithic Quantity (p = 0.004) and Flaking Inefficiency (p = 0.049). For both Quantity and Flaking Inefficiency, participant age was the only coefficient that contributed to its respective model with significance, such that for Quantity, p = 0.040; and for Flaking Inefficiency, p = 0.036.

Because the male and female participant groups contain both adults and children, who have already demonstrated differences in grip strength, MRT score, and Fitts' Law Test response times, it is likely that the significant values returned by the models conducted on the basis of participant sex are driven by the effects of age, and, consequently, of maturation, on knapper ability.

WORKS CITED

Benet-Martínez, Verónica, and Oliver P. John. 1998. Los Cinco Grandes across cultures and ethnic groups: Multitrait-multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology* 75.

Busato, Vittorio V., Frans J. Prins, Jan J. Elshout, and Christiaan Hamaker. 1999. The relation between learning styles, the Big Five personality traits and achievement motivation in higher education. *Personality and Individual Differences* 26:129-140.

Caprara, G. V., M. Vecchione, G. Alessandri, M. Gerbino, and C. Barbaranelli. 2011. The contribution of personality traits and self-efficacy beliefs to academic achievement: a longitudinal study. *Br J Educ Psychol* 81 (Pt 1):78-96.

Cohen, Jacob. 1962. The Statistical Power of Abnormal-Social Psychological Research: A Review. *Journal of Abnormal and Social Psychology* 65:145-153.

Delagnes, A., and H. Roche. 2005. Late Pliocene hominid knapping skills: the case of Lokalalei 2C, West Turkana, Kenya. *J Hum Evol* 48 (5):435-72.

Fitts, Paul M. 1954. The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement. *Journal of Experimental Psychology* 47:381-391.

Guadagnoli, Edward, and Wayne F. Velicer. 1988. Relation of Sample Size to the Stability of Component Patterns. *Psychological Bulletin* 103:265-275.

Kline, Michelle Ann. 2016. TEACH: An Ethogram-based Method to Observe and Record Teaching Behavior. *Field Methods* 29 (3):205-220.

Morgan, T. J., N. T. Uomini, L. E. Rendell, L. Chouinard-Thuly, S. E. Street, H. M. Lewis, C. P. Cross, C. Evans, R. Kearney, I. de la Torre, A. Whiten, and K. N. Laland. 2015.

Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nat Commun* 6:6029.

Neu, C.M., F. Rauch, J. Rittweger, F. Manz, and E. Schoenau. 2002. Influence of puberty on muscle development at the forearm. *American Journal of Physiology - Endochronology and Metabolism* 283:E103-E107.

Pargeter, J., N. Khreisheh, and D. Stout. 2019. Understanding stone tool-making skill acquisition: Experimental methods and evolutionary implications. *J Hum Evol* 133:146-166.

Semaw, S., P. Renne, J.W.K. Harris, C.S. Feibel, R.L. Bernor, N. Fesseha, and K. Mowbray. 1997. 2.5-million-year-old stone tools from Gona, Ethiopia. *Nature* 385:333-336.

Soto, C. J., O. P. John, S. D. Gosling, and J. Potter. 2011. Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample. *J Pers Soc Psychol* 100 (2):330-48.

Stout, D., J. Quade, S. Semaw, M. J. Rogers, and N. E. Levin. 2005. Raw material selectivity of the earliest stone toolmakers at Gona, Afar, Ethiopia. *J Hum Evol* 48 (4):365-80.

Stout, Dietrich, E. Hecht, N. Khreisheh, B. Bradley, and T. Chaminade. 2015. Cognitive demands of lower paleolithic toolmaking. *PLoS One* 10 (4):e0121804.

Stroop, J.R. 1935. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology* 18:643-662.

Vandenberg, Steven G., and Allan R. Kuse. 1978. Mental Rotations, A Group Test of Three-Dimensional Spatial Visualization. *Perceptual and Motor Skills* 47:599-604.

Wind, A. E., T. Takken, P. J. Helders, and R. H. Engelbert. 2010. Is grip strength a predictor for total muscle strength in healthy children, adolescents, and young adults? *Eur J Pediatr* 169 (3):281-7.

Wind, Anne E., Tim Takken, Paul J.M. Helders, and Raoul H.H. Engelbert. 2010. Is grip strength a predictor for total muscle strength in healthy children, adolescents, and young adults? *Eur J Pediatr*.

CHAPTER 3: DISCUSSION AND CONCLUSION

DISCUSSION

Questions of whether – and, more recently, of how – children contributed to the material culture of prehistory have existed in the archaeological literature for decades (Lillehammer 1989). Although great strides have been taken to address this question regarding children from relatively more recent time periods (Baxter 2022; Derevensk 2000; Nowell 2021), it is still largely unknown how and to what extent Paleolithic children engaged in one of the earliest markers of hominin material culture: stone toolmaking. Much remains unclear, from the process of identifying novice knappers in the archaeological record (Gamble and Porr 2005) to understanding who these novice knappers were (Ferguson 2008; Högberg 2018) – were they children, and if so, at what age did they start acquiring stone toolmaking skills? Or were they adults, as in the case of modern examples of stone knapping apprentices (Roux, Bril, and Dietrich 1995; Stout 2002)?

Within the answer to this question lies a greater understanding of what sort of social learning processes would have been advantageous, or even necessary, for the transmission of skilled stone toolmaking knowledge, given that individuals at different developmental stages may require different methods of teaching to become successful at their task. In other words, children may learn differently than adults, and because of this, the selection pressures related to stone toolmaking skill acquisition may be different for children than from what experiments with modern adult participants conclude. For example, in research on language learning across the lifespan, the concept of a ‘critical period’ is often cited as a distinct feature of learning in infancy through early childhood (Snow and Hoefnagel-Höhle 1978). This particular aspect of cognitive flexibility is said to wane as individuals age, though this claim is contested (Scovel 2003), thus making second language learning purportedly more difficult during adulthood. A similar ‘golden period’ has been described for motor learning, ranging from 6 – 12 years old (Hirtz and Starosta 2002), though, as with language acquisition, the presence and influence of such a learning period is unconfirmed (Solum, Loras, and Pedersen 2020).

If there is an equivalent ‘critical period’ for skill acquisition, the learning needs of children in a skills-based task such as stone toolmaking may not resemble those of adult participants in modern experimental settings. However, even if the performance differences between adults and children are a result of their differing cognitive and motor maturation, as some researchers suggest (Lukacs and Kemeny 2015), and not due to a form of ‘implicit learning’ that promotes child learning and may give them a paradoxical advantage over novice adults (Reber 1989), this developmental gulf would still yield a learning landscape where multiple forms of instruction emerge to the benefit of learners of different ages. With this in mind, the outstanding question is not *whether* adults and children learn differently, but rather *how* their learning is different and whether children find ways to compensate for their relative lack of cognitive and motor affordances to achieve similar outcomes to those of adults in a skills-based task.

The effect of other unknowns, such as the total amount of training time and the degree of practice density required to gain proficiency, may also play a role in the outcomes of novice knappers of different ages. Although the present study cannot and does not attempt to address all these questions, it is important to establish a baseline for future lines of inquiry on this topic: namely, are there measurable differences in the lithic products produced by child and adult knappers instructed under the same conditions? The results of this research indicate that there are.

Knapper Outcomes: The Effects of Motor and Cognitive Ability and Personality on Quantity, Quality, and Flaking Inefficiency

Motor Ability

Perhaps the most surprising result of this study is the distinct performance of the adult male participants. Although adult participants of both sexes outperformed children both in terms of their knapper skill metrics and psychometric and motor testing, the Quality of lithics produced by adult male participants far surpassed that of any other experimental subgroup. When comparing adult male and adult female participants' psychometric and motor performances, the only measure that yielded a significant difference between these groups was grip strength, where males were, on average, stronger than females. Grip strength positively correlated with lithic Quality in children and in male participants across conditions and negatively correlated with Flaking Inefficiency in both males, generally, and specifically in child males, indicating that male participants with higher grip strength were more likely to efficiently remove mass from their cores. Both trends indicate a relationship between grip strength and lithic Quality.

Although grip strength was the most reliable predictor of Quality, it had few correlations with Quantity, somewhat in contrast to previous findings that note the connection between knapper productivity and grip strength (Pargeter et al. 2023). Female participants did demonstrate a relationship between grip strength and Quantity, a statistic likely driven by the adult women, who produced more lithics than any other experimental subgroup, and given that boys produced more than girls among children. However, because adult female and adult male participants did not statically differ in lithic Quantity, it is unlikely that grip strength is strongly correlated with Quantity in this sample.

Motor accuracy was assessed in this study using a digital version of a Fitts' Law reciprocal tapping task, which examines the trade-off in human movement between speed and accuracy (Fitts 1954). The mastery of this trade-off has been hypothesized to be relevant to knapping skill acquisition (Pargeter et al. 2023; Stout 2002) as it helps knappers manage the accurate application of force during stone toolmaking. In this study, lower values for Fitts' Law Test (FLT) response times negatively correlated with lithic Quantity and Quality, meaning that participants who responded more quickly to the FLT task generated more lithic products of better quality. Specifically, speed negatively correlated with Quantity in children, male participants, female participants, and child males. Speed also negatively correlated with Quality in male participants, and it positively correlated with Flaking Inefficiency in both male and female participants, such that participants with slower response times were also more likely to ineffectively remove mass from their cores.

FLT accuracy scores produced few correlations with knapper outcomes: in female participants, accuracy scores negatively correlated with Quantity, and in male participants, they negatively correlated with Quality and positively with Flaking Inefficiency. Because the FLT accuracy score is calculated as a percent error, lower accuracy scores are indicative of better performance at the task, and as such, the negative correlations with Quantity and Quality indicate that participants who responded more accurately to the task were also more likely to produce more and better lithics, whereas the positive correlation with Flaking Inefficiency demonstrates that participants who were less accurate were also less efficient at removing stone mass from their cores.

Given that both measures of participant motor ability were predictive of favorable outcomes with knapper skill performance, I wanted to explore the possibility that grip strength and FLT scores correlate with each other among participants. If a positive correlation were to exist, this may be indicative of an overall effect of participant fitness, or 'athleticism,' on knapper ability. To do this, I first examined the relationship between the two FLT metrics and found that, when all participants were tested together, FLT response times and accuracy scores demonstrated a weak, but significant positive correlation ($r(45) = 0.335$, $p = 0.021$). When adults were tested separately, this positive correlation became negative, though not at significance ($r(28) = -0.310$, $p = 0.096$), which indicates, marginally, that adults who committed fewer errors during the FLT task also had slower response times and thus demonstrated a classic Fitts' Law speed-accuracy trade-off (*Figure 3.1*). Children showed no relationship between speed and accuracy ($r(16) = 0.080$, $p = 0.753$). This difference between developmental groups is consistent with adults' better performance at the FLT task.

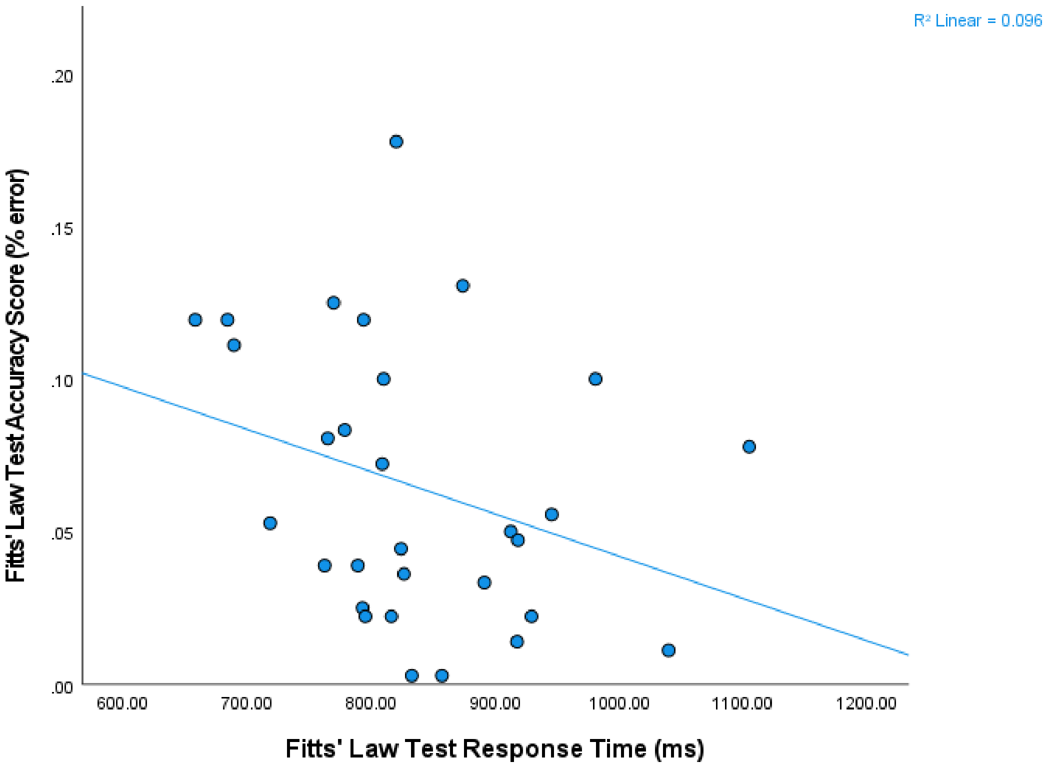


Figure 3.56: Adult Participant Performance in Fitts' Law Test. Adult participants with higher FLT accuracy scores also performed the task more slowly, thus demonstrating a classic Fitts' Law speed-accuracy trade-off.

To examine the relationship between FLT performance and grip strength, I created an interaction term for FLT response times and accuracy scores to represent overall aptitude at the FLT task. Across all participants, FLT aptitude negatively correlated with grip strength (r(43) = -0.324, p = 0.030) at significance. Because lower FLT aptitude scores are indicative of better performance, given that shorter response times and fewer errors generate smaller values, this negative correlation suggests that participants who demonstrated higher grip strengths were also more likely to achieve better FLT scores. When tested separately, this correlation disappeared both in adults (r(27) = 0.038, p = 0.843) and in children (r(14) = -0.176, p = 0.515). When tested by participant sex, there was no significant correlation between FLT aptitude and grip strength in male participants, as a whole (r(17) = -0.352, p = 0.139) and when tested by developmental group (Adults: r(6) = 0.493, p = 0.214; Children: r(9) = -0.160, p = 0.637). However, there was a significant negative correlation within female participants (r(23) = -0.441, p = 0.027) (*Figure 3.2*). To test whether this significant relationship was a byproduct of differences in participant ages, I examined female adults and children separately and found that within developmental groups, a correlation between grip strength and FLT aptitude also fell away (Adults: r(18) = 0.001, p = 0.998; Children: r(3) = -0.516, p = 0.373).

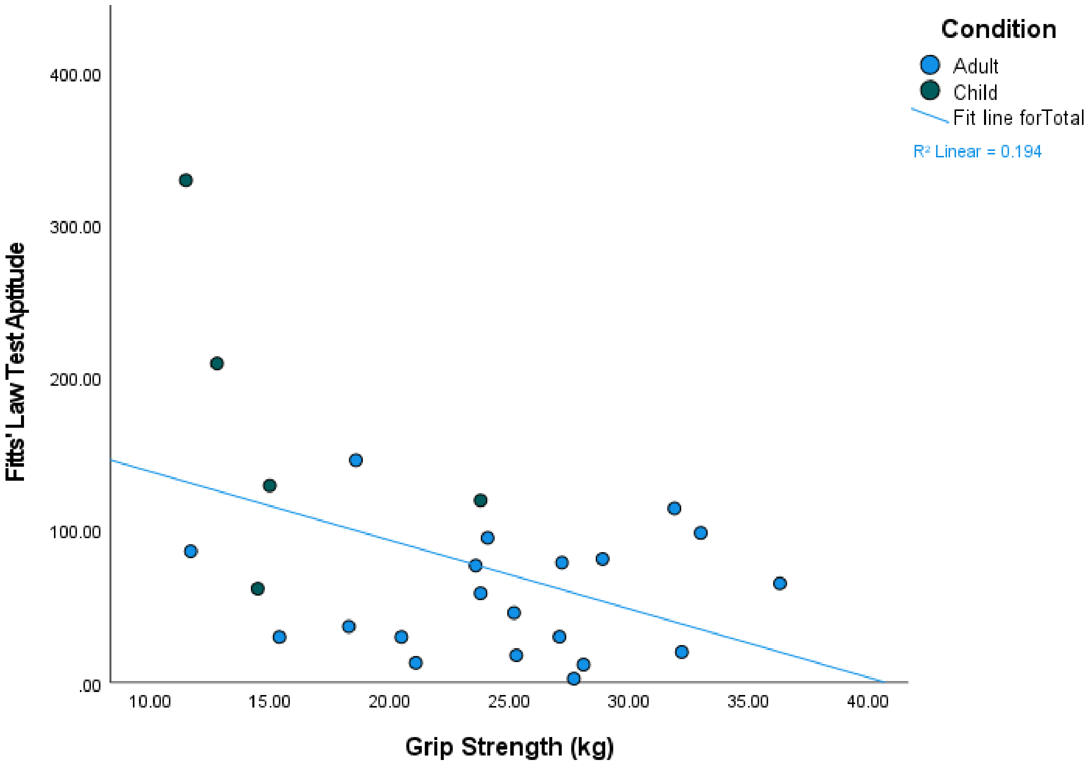


Figure 3.57: Correlation of Grip Strength with Fitts' Law Test Aptitude in Female Participants. Female participants, generally, experienced a relationship between grip strength and FLT aptitude, such that stronger women and girls were also more proficient at the FLT task.

The significant relationship between FLT aptitude and grip strength demonstrated across all participants suggests that, in general, greater participant athleticism is predictive of knapper success. However, given that this trend disappears within participant subgroups, it is likely that it is a byproduct of participant age and maturational differences between developmental conditions. In other words, just as grip strength and FLT aptitude increased with age, so did the influence of motor ability on knapper performance.

Cognition

While each of the psychometrics used in the study were validated for both developmental groups, adults routinely outperformed children in the majority of psychometric tasks, with the exception of the Stroop Color-Word Test and the BFI, where they showed no difference. This is likely due to one of several factors. First, and most probable, the adults demonstrated better psychometric performances because they were more developmentally mature in the tested domains (Kail, Pellegrino, and Carter 1980; Kerr 1975; Luna et al. 2004). Second, it is possible that the adults outperformed the children because they already had some familiarity with the format of the psychometrics, themselves (Cherney and Neff 2004), either through previous coursework (e.g. an introductory psychology class) or due to the popularity of psychometric-like online test-taking (e.g. online quizzes designed to ‘test your intelligence’ through media such as *Buzzfeed*). Finally, some of the children expressed dismay when asked to complete the psychometric tasks, as this defied their expectations of what should constitute a summer program setting (i.e. it was ‘too much like school’). While most children appeared sincere in their efforts to complete the tasks, some may have been distracted and therefore would not have performed at their best.

Each of the psychometrics selected for this study were chosen because they measured an aspect of cognition hypothesized to promote successful stone toolmaking. The MRT, aptly named to examine individuals' ability to mentally manipulate objects (Vandenberg and Kuse 1978), was used in this study to evaluate participants' spatial cognition, which aids knappers in their ability to keep track of relevant core morphological features, such as viable platform angles (Stout et al. 2000; Wynn 2002; Wynn and Coolidge 2016). Sex-based performance differences in this task have long been reported, favoring males (Parsons et al. 2004), though there is evidence to suggest this effect may be mediated by the gendered nature of spatial activities children are exposed to as they age (Nazareth, Herrera, and Pruden 2013), the confidence of the individual performing the task (Estes and Felker 2012), and educational experience (Richardson 1994). Although this study reported no sex-based differences in MRT performance, male participants' MRT scores did correlate more frequently with knapper outcomes than did those of female participants.

To review, the MRT positively correlated with both Quantity and Quality and negatively with Flaking Inefficiency in male participants across conditions. It also positively correlated with Quantity and negatively with Flaking Inefficiency in female participants. When tested within developmental groups, these trends fell away for female participants but persisted in Quantity for male children. These data do suggest that the novice knappers with higher MRT scores saw some benefits to the Quantity of lithics they produced, an effect that may become more pronounced as individuals gain expertise and begin approaching the knapping task in cognitive chunks, as opposed to individual action sequences (Leone et al. 2014; Moore 2020). The negative correlation with Flaking Inefficiency indicates that participants who performed better at the MRT task were also more capable of effectively reducing their cores. This result is logical, given that a novice knapper who is proficient in mentally tracking core morphology would also be more proficient in exploiting it, thereby producing larger and more useful debitage, as opposed to shatter.

The Stroop Raw Color-Word Test, designed to evaluate selective attention and processing speed (Stroop 1935), was used in this study to examine participants' inhibition control, a feature of executive functioning hypothesized to promote successful stone toolmaking (Putt, Wijeakumar, and Spencer 2019; Hecht et al. 2023). Participant Stroop scores yielded very few correlations with knapper outcomes. In adults, there was a marginal (p = 0.06) positive correlation between Stroop score and Quantity, and in female participants, scores positively correlated with Flaking Inefficiency. The effect sizes of both these correlations were weak-to-moderate, at r(28) = 0.347 and r(27) = 0.442, respectively. The Stroop Test was also the only psychometric that demonstrated no statistically significant differences between any of the experimental groups. With this shortage of significant results, it seems clear that inhibition control did little to influence knapper outcomes in this study.

This result, while unexpected, especially within child participants, whose inhibition control continues to increase into young adulthood (Williams et al. 1999), is consistent with the literature. Previous studies of stone toolmaking that emphasize the importance of executive functioning often relate it to Acheulean technology, which requires more sophisticated use of action planning (Stout et al. 2008). While the degree to which advanced cognitive ability plays a role in Oldowan-style knapping remains an open question (Wynn et al. 2011), its influence should not be discounted, completely, given the following: By the advent of stone tool technology in the archaeological record, ancient hominin knappers had already diverged from the last common ancestor with extant nonhuman apes and had, therefore, already undergone some measure of modern-humanlike cognitive evolution (Toth and Schick 2018). Given the correlation in this study between MRT performance, which is an indicator of working memory capacity, and knapper success, it is evident that cognition does play some role in the production of Oldowan-style stone tools, even if its influence is not as prominent as in Acheulean-style technology.

Personality

The influence of personality in stone toolmaking success has not been widely studied. In this dissertation, it was hypothesized that participants with higher Openness and Conscientiousness scores, as measured by the Big Five Inventory (BFI) would perform better at the stone toolmaking task than their peers with lower scores in these domains, given that both have been demonstrated to correlate with intrinsic motivation and academic success in children (Caprara et al. 2011) and young adults (Busato et al. 1999; Komarraju, Karau, and Schmeck 2009). The BFI offers one of the most validated measures of personality across an individual's stages of development, and self-reports on the Big Five dimensions significantly self-correlate starting as early as 5-years-old (Measelle et al. 2005). Although individuals do experience changes in personality across their lifetimes (Srivastava et al. 2003), these changes are typically predictable and in accordance with age-appropriate trends (Soto et al. 2011). Namely, Openness, Conscientiousness, and Agreeableness are all expected to increase as individuals age, whereas Neuroticism declines, and Extroversion levels remain approximately the same.

Participants in this study showed no difference in personality across developmental groups, although differences did emerge between male and female participants, with the former reporting higher Openness scores and the latter displaying higher scores in Neuroticism. In adults, generally, and adult males, specifically, higher scores in Agreeableness negatively correlated with Flaking Inefficiency, indicating that those who were more Agreeable were also more likely to effectively reduce their cores while minimizing shatter. In male children, Openness, Conscientiousness, and Agreeableness all negatively correlated with lithic Quantity, while in female children, high Neuroticism scores positively correlated with Quantity. None of the BFI traits correlated with knapper outcomes in adult females, whereas in adult males, Conscientiousness positively correlated with Quantity, and Agreeableness negatively correlated with Flaking Inefficiency.

Within each of the BFI domains are what have been described in the literature as particular facets of that trait (Costa Jr. and McCrae 1995). Namely, Openness is often subdivided into Openness to Ideas and Openness to Aesthetics, Conscientiousness into Self-Discipline and Order, Extroversion into Activity and Assertiveness, Agreeableness into Altruism and Compliance, and Neuroticism into Anxiety and Depression. Although this study did not employ a version of the BFI that examined facets of the primary domains, I will briefly consider them here as a means of interpreting the correlations of personality and Quantity in children.

The following are anecdotal and based on observations I made while working with the child participants. Specifically, I observed that the boys who repeatedly struck their cores in rapid succession and with little regard to where or how they were percussing were more likely to successfully detach lithic pieces compared to their more deliberate, fastidious peers. This sort of 'wild abandonment' approach to knapping (one congruent with lower scores in Openness: Openness to Ideas, Conscientiousness: Self-Discipline and Order, and Agreeableness: Compliance) may have ultimately increased their Quantity output because, regardless of technique, they may have introduced enough incipit fractures into the core that pieces eventually came off. Conversely, the girls generally approached knapping more timidly, often citing concerns about hitting their fingers with the hammerstone. Their higher Neuroticism scores may have resulted in a more methodical approach to knapping, one that caused them to more carefully consider and make use of the demonstrated knapping techniques, which may have resulted in an increase in their lithic Quantity.

Converse to male children, adult males saw a positive correlation between Conscientiousness and lithic Quantity, a trend that is in line with the increase in Conscientiousness across an individual's lifetime (Soto et al. 2011). This change indicates that as male participants age, they may shift their knapping approach from one that is bombastic to one that is more systematic and intentional. The negative correlation in adult male participants between Agreeableness and Flaking Inefficiency suggests that adult males who were more Agreeable, and by extension more Compliant, may be those who followed the advice given to them during instruction and subsequent demonstrations, therefore increasing their chances of successful flake detachments and more efficiency core reduction.

The impact of personality on knapper outcomes is still poorly understood, and future studies on this subject would benefit not only from a longitudinal design, but also from the inclusion of behavioral data. Such data were collected during this dissertation, in the form of videos, and to statistically ground the speculations made in the above paragraphs, it will be important to code and analyze those data, a project that will be embarked upon in future endeavors.

Summary

In all, metrics of motor ability and proficiency in the MRT were the strongest predictors of knapping success in this study. Given that children performed more poorly than adults in each of these key tasks (i.e. grip strength, FLT aptitude, and MRT scores), it follows that the values of their knapper skill metrics were correspondingly less favorable. These results imply a certain baseline muscle strength and coordination, in addition to a basic understanding of how to keep track of and manipulate core morphology, is required to produce useful Oldowan-style stone tools. The correlation between grip strength and lithic Quality is especially evident in this sample, given that adult male participants possessed both the highest grip strength scores and produced the best Quality lithics. Because adult female participants' MRT and Fitts' Law Test scores were equivalent to those of their adult male counterparts, the latter subgroup's greater grip strength provides the best explanation for why adult male participants produced lithics of much better Quality than adult females. Consequently, because adult female participants produced the greatest Quantity of lithics, it would seem that even if individuals are capable of exploiting a core's fracture mechanics to successfully create lithic products, irrespective of utility, if a minimum grip strength prerequisite is not met, cognitive ability appears to not be enough to compensate for what a knapper lacks in overall fitness, at the novice level. In other words, possessing a higher grip strength may allow novice knappers to create useful lithic products earlier in their skill acquisition process than their weaker counterparts, given that grip strength has been demonstrated to be relevant to control and support of the hammerstone (Williams-Hatala et al. 2018) and the core (Faisal et al. 2010; Key and Dunmore 2015) and aids in generating kinetic energy for flake removal (Nonaka, Bril, and Rein 2010).

As knappers become more expert, they exert less pressure overall on their hammerstone, and the localization of non-thumb digit pressure shifts from digits III to V (the middle, ring, and pinky fingers) to digit II (the forefinger), with both novice and expert knappers making regular use of digit I (the thumb) (Williams-Hatala et al. 2021). Such findings provide evidence that knappers revise their hammerstone use technique as they become more skilled and that grip strength may play less of a role in successful flake removals, as a consequence, given that experts exert less pressure on their hammerstones than novices. If this is the case, one may expect to see novice knappers with lower grip strengths ‘catch up’ to their stronger peers in terms of lithic Quality as both groups refine their knapping knowledge and technique. Should such an equalization take place, differential performance on psychometric tasks and personality testing may become more predictive of knapper outcomes in participants at later stages of learning.

Learning Effects

Of the three knapper outcome variables, only Quantity demonstrated a learning effect over participants’ five days of training. There are several, likely overlapping, explanations that could account for this trend. First, participants may have gradually become more comfortable with applying the force necessary to detached pieces from the core, as they became more familiar with how to use the hammerstone effectively and how to position their fingers around the core so that they were not at risk of injury. Second, participants may have started to recognize the salient aspects of core morphology, per the effects of experience and expertise on pattern recognition and chunking (Guida et al. 2012; Pargeter, Khreisheh, and Stout 2019; Geribàs, Mosquera, and Vergès 2010; Dreyfus 2016), and may have begun to exploit them, resulting in more successful detachments, regardless of lithic product Quality. Finally, by the end of five days of practice, participants may have additionally started to internalize knapping advice related to body kinematics and manual gestures, specifically involving upper-body posture and hammerstone trajectory, velocity, angle, and grip (Bril et al. 2010; Li et al. 2022; Vernooij, Mouton, and Bongers 2012; Williams-Hatala et al. 2021). To test these interpretations, future studies of this nature should include entry and exit surveys that assess: 1) participants' level of comfort and familiarity with knapping procedures (e.g. how safe they perceive stone toolmaking to be and their knowledge of aspects of knapping such as where to position their fingers around the hammerstone and core), 2) participants' understanding of and ability to describe the task goals (e.g. how to identify a 'useful' whole flake, the steps necessary to detach flakes from the core, and how to recognize a good place to strike the core), and 3) a self-assessment on how their own knapping skills and knowledge have changed over the course of training.

Interestingly, adult male participants were the only subgroup that, when tested individually, did not maintain a statistically significant increase in lithic output (Quantity) over the training period, once again making them the odd-group-out. As with the trends displayed in effects of motor ability and psychometric scores on knapper outcomes, this is likely due to their greater grip strength giving them an initial advantage as novice-stage stone toolmakers. In addition to giving them an early edge in lithic Quality, adult males' higher grip strength scores may have caused them to begin producing more lithic products sooner in their training than other participants, which would have made their rate of improvement more subtle and under the threshold of statistical significance.

The lack of learning effects in Quality and Flaking Inefficiency are likely attributable to the study's short training times, discussed in greater detail in §Limitations. Although 2.5 hours of knapping practice was enough to see an improvement in the Quantity of participants' lithic output, it was not long enough for participants to improve the Quality of their flakes. In fact, the Quality of participants' lithics actually declined over their five days of practice, which may be indicative of a threshold in their learning process whereby they were starting to grasp the *conaissance* (i.e. cognitive understanding) of stone toolmaking without also having developed its *savoir-faire* (i.e. motor ability and know-how) (Bamforth and Finlay 2008). The former can be explained in an afternoon, and with regular practice and consistent reminders about proper knapping principles from a more knowledgeable individual, novices seem to pick up on the basics of fracture mechanics and whole flake production rather quickly. The latter, however, takes much longer to acquire, as the embodiment of knowledge requires the gradual process of enskilment (Ingold 2001). Differences in lithic quality between knappers of varying levels of expertise can be felt even after years of apprenticeship (Roux, Bril, and Dietrich 1995), and although it is unlikely that mastery of an Oldowan-style stone toolmaking task requires as much time and practice as later stone industries (e.g. Acheulean, Levallois), it is clear that the amount of training provided in this study was not adequate for participants to even improve the Quality of their flakes, let alone obtain toolmaking proficiency.

The influence of individual differences in psychometric and motor ability on participant learning were most visible at the level of participant subgroups. In adults, generally, and particularly in adult female participants, those with faster FLT response times demonstrated larger learning effects for lithic Quantity (i.e. ΔQuantity), meaning that better FLT performance was tied to an improvement in lithic output over the duration of the experiment. Although the relationship between speed and ΔQuantity was present in all adult participants ($p = 0.045$), it was especially prominent in adult women ($p = 0.022$), who produced the most lithics of any participant subgroup. Just as grip strength was predictive of lithic Quality in adult men, it may be that FLT response times are responsible for the greater Quantity of lithics produced by adult women. If this is the case, adult female participants may have compensated for their relative lack of upper body strength, as compared to the adult males, with a strategy similar to that of the male children; namely, that they struck their cores more rapidly and with higher frequency in an attempt to remove a flake from a particular platform. In other words, they may have ‘bashed’ their cores because they either did not possess or did not properly employ adequate grip strength to remove the flake in a single strike. This resulted in the production of many, poor quality detached pieces, likely due to the introduction of incipit fractures in the core through repeated, unsuccessful strikes to the same platform and the consequent ‘rounding’ of the platform in question.

Interestingly, MRT scores negatively correlated with ΔQuality in adults; however, this trend is consistent with the overall decline of lithic Quality during the training period. MRT scores positively correlated with Quantity and negatively with Flaking Inefficiency across participant subgroups; however, MRT performance did not demonstrate any significant correlations with Quality. It is possible that the ability to mentally track objects has a greater influence on a knapper’s potential to remove mass from a core, broadly and irrespective of the quality of lithics produced, given that this cognitive capacity is especially relevant to identifying viable striking platforms. In other words, even if a participant is capable of knowing where to strike a core to generate a lithic product, this may not be enough to guarantee that the resulting product will be useful. With that said, it is also possible that this negative result is related to the experiment’s relatively short training times and that MRT scores may be positively predictive of knapper success as participants see increases in the overall Quality of their lithics. Among children, MRT scores were positively correlated with increases in lithic Quantity. The absence of this same correlation in adults indicates that children, who have less developed motor abilities, as evidenced by their lower grip strength and FLT aptitude values, may have relied on their cognitive engagement with the task to compensate for their relative lack of fitness.

Any correlation between personality and learning effects of knapper outcomes, as with all inferences on the relationship between personality and knapping ability that this study offers, should be taken as exploratory, as future behavioral work on the topic is necessary to provide an empirical backing for the following observations. More Agreeable children, male participants, and, particularly, male children, were both more likely to see increases in their lithic Quality during the training period, which may be attributable to their willingness to receive assistance with the task and to follow the instructions provided during the task demonstration.

Conversely, Agreeableness was negatively correlated with improvements in Quality among adult female participants, such that the more Agreeable women produced less useful lithics by the end of their training period. Because low Agreeableness scores are typically associated with critical attitudes, skepticism, and a general disregard for maintaining positive relationships with other people (Costa Jr. and McCrae 1995), the negative correlation between Agreeableness and improvements in lithic Quality indicates that the adult female participants who tend to be less socially invested were more likely to produce useful stone tools. Adult female participants also demonstrated a negative correlation between Openness and ΔQuantity, which, in combination with the negative association between Agreeableness and ΔQuality, paints a rather unflattering picture of successful female knappers. It seems premature and potentially harmful to conclude that less pleasant personalities in women are predictive of more effective knapping skill acquisition. Instead, it is possible that these results are capturing the phenomena that women, who typically score higher in Agreeableness than men (Gensowski, Gortz, and Schurer 2021; Schmitt et al. 2008), are thus more likely to be ‘people-pleasers,’ which may, paradoxically, make them less inclined to ask for assistance when it is needed. If such was the case in this study's sample, women who were more Agreeable may not have solicited the help required to see improvements in their lithic Quality over the experiment's training period.

The negative correlation between Openness and ΔQuantity in adult female participants is somewhat more difficult to explain. Given that all participants were naïve to stone toolmaking, knapping would constitute a new experience for participants, and higher Openness scores should indicate individuals' willingness to try new things. Low Openness scores imply that an individual is resistant to new experiences and other people's ideas, which, in the context of the experiment, may manifest as a reluctance to follow task instructions. As with Agreeableness, adult female participants with higher Openness scores may have been paradoxically disadvantaged in the task, as they may have tried to adhere to the instructor's prescribed knapping technique at the detriment of their lithic productivity. In other words, these participants may have been so intent on knapping 'correctly' that they hindered their own progress, within the timeframe of the experiment. This interpretation is corroborated by female participants' higher Neuroticism scores. Those with lower Openness scores were also the individuals who reported faster FLT response times, at marginal significance ($r(19) = 0.427$, $p = 0.053$) (*Figure 3.3*), which supports the previously discussed finding that adult female participants who produced larger quantities of lithics likely did so through a bombastic style of knapping: one that emphasized quick, repetitive strikes, counter to the methods advised by the instructor (e.g. deliberate, planned strikes on viable platforms).

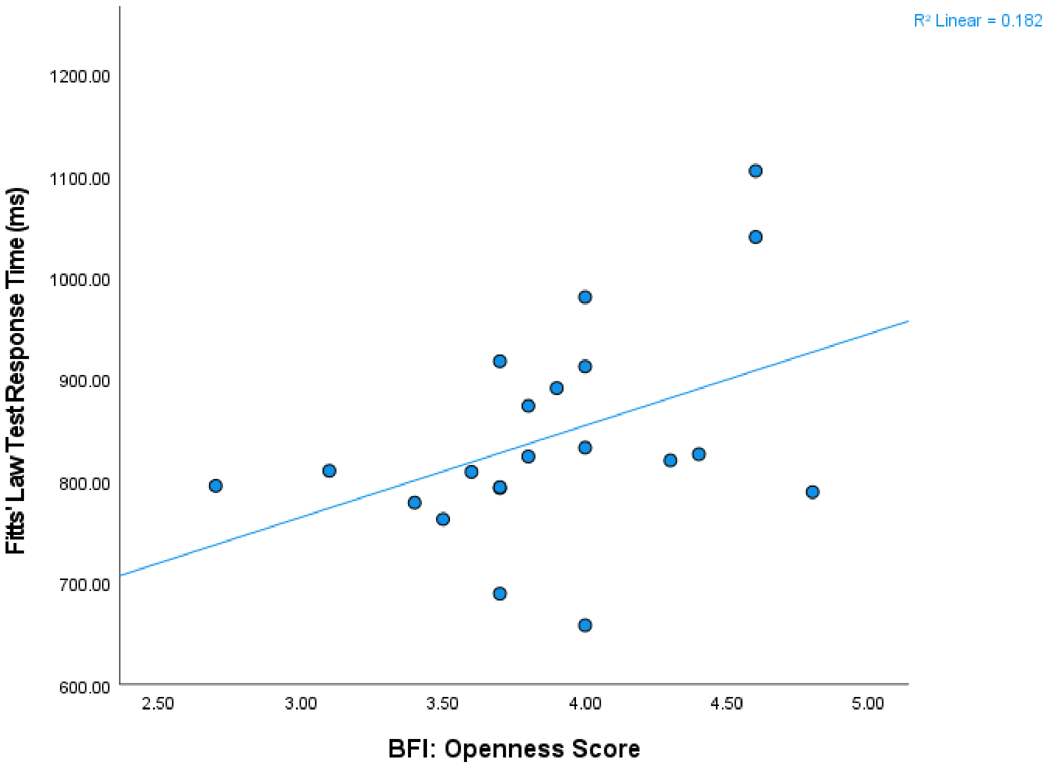


Figure 58.3: Correlation of BFI: Openness Score with Fitts' Law Test Response Time in Adult Female Participants. Adult female participants with lower Openness scores were also more likely to respond more quickly to the FLT task.

Adult male participants who reported higher Conscientiousness scores were more likely to experience a negative correlation with ΔFlaking Inefficiency, such that more conscientious men saw improvements in efficient removal of mass from their cores. This relationship is a logical one, as Conscientiousness encompasses the domains of Order and Self-Discipline. Successful knapping requires the careful application of striking precision and velocity, and when executed in a disciplined manner, it is more likely that a knapper will produce useful lithics and, by extension, more efficiently reduce their cores. Among female children, Conscientiousness was positively correlated with ΔQuantity, likely for similar reasons. Namely, a more deliberate application of knapping techniques should result in a greater number of lithic products, regardless of their overall Quality, and more Conscientious individuals should demonstrate the grit required to make improvements in their output as their knapping skills increased with practice.

Broader Implications

Colloquially, childhood is described as a time when individuals experience ‘critical periods’ of skill-based learning, whether these skills are related to language learning (Snow and Hoefnagel-Höhle 1978), motor ability (Hirtz and Starosta 2002), or other domains of knowledge. However, this ‘critical period’ model of learning has been contested (Scovel 2003; Lukacs and Kemeny 2015; Solum, Loras, and Pedersen 2020), and the results of this study indicate that this challenge to common knowledge perspectives of childhood skill acquisition is warranted. Namely, child participants, who all fell within the middle childhood or early adolescence age brackets (8- to 13-years-old) did not achieve comparable knapping outcomes with their adult counterparts. Instead, their performance was poorer across knapper skill metrics (Quantity, Quality, and Flaking Inefficiency) and within psychometric, motor, and personality testing, except in the cases where no statistical differences were present (i.e. Stroop Color-Word Test).

Despite the binary treatment of children and adults in this study, it was evident that the effects of age on task performance were more linear in nature, with younger children demonstrating the least developed abilities and older children approaching adult aptitudes. This was the case with grip strength, MRT scores, and FLT performance, all of which were strongly predictive of knapper skill. These results indicate that improvements in stone knapping ability are likely tied to the maturation of cognition and motor function and that children’s attempts at stone toolmaking will not match adult performance until these developmental milestones have been met. That being said, this study provides a snapshot of the stone toolmaking learning process, one that favors direct-active teaching as its primary mode of skill transmission. It is likely that in communities where individuals made and used stone tools routinely to complete everyday tasks, children would have been exposed to stone knapping activities from an early age. Such opportunities for observational learning would almost certainly have impacted their knowledge of the mechanics of stone toolmaking, which could have accelerated their ability to produce useful stone tools once they engaged actively in the toolmaking process, themselves. Because of this difference in learning environments, and due to other factors including the disparate motivational context in which the stone tools are being made and the types and frequency with which individuals engage in gross motor activity, it is important to keep in perspective that the knapping performance of modern human children should not be considered equivalent to that of ancient hominin children of a similar age and stage of development.

It is likely, however, that ancient hominin children would have also needed to mature to a particular threshold of cognitive and motor ability to become successful stone toolmakers, as the physics of stone fracture mechanics have not changed in the interim millennia. The age at which this threshold would have been reached may differ from that of modern humans, given differences in the early toolmakers' hand morphology and cognitive capacity.

Although these data cannot definitively address the question of novice stone toolmaker identity in the archaeological record, ethnographic evidence of the acquisition of skilled behavior among extant knapping communities and forager societies suggests that children may be orthogonally involved in skilled activities, as their bodily affordances and cultural norms allow. For example, Bird and Bliege Bird (2000; 2002) describe the involvement of children as young as 5-year-old in coastal foraging among the Meriam of the Torres Straight Islands. At this age, children’s contribution to the foraging efforts is minimal, as they primarily tag along with older children and adults, who help them identify prey items and differentiate between species of varying value. The older children who engage in foraging are limited in their efficacy, due to their shorter stride length, which limits the quantity of prey items they can acquire in the same time as adult foragers, and grip strength, which impacts their ability to process harder to open shellfish. However, even though children are less capable of achieving the same foraging yield as adults, Bird and Bliege Bird mention that the children possess the requisite knowledge to successfully forage high value prey items but lack the bodily affordances to perform at the level of adult competence.

The ethnographic literature on living stone knapping communities of practice is quite sparse, as the regular practice of stone knapping has almost entirely been replaced by more recent technological innovations. The studies that do exist on this topic all describe novice or apprentice knappers as adolescents (Weedman Arthur 2010) or adults (Stout 2002; Roux, Bril, and Dietrich 1995), who undergo a training period of at least five years before they are considered competent enough to work with good quality raw materials. However, although active training in stone knapping is delayed until these older ages, two of these ethnographies describe the peripheral involvement of children in stone toolmaking activities.

In the case of the adze makers of Langda village in Indonesian Irian Jaya, adult male knappers of varying skill levels would travel together to acquire raw stone materials from the nearby Ey River Valley (Stout 2002). Children of both sexes were permitted to accompany the men on these trips, although the boys were given opportunity to assist with the work when extra help was needed to process especially large boulder cores. Weedman Arthur (2010) goes into even greater detail about the scaffolded learning process among the Konso women hide-workers of southern Ethiopia. Within this community of practice, girls begin learning the craft of hide-working around the ages of 6 to 8 years old. At this stage, the girls do not work with the stone materials, themselves, but instead are taught other aspects of the craft, such as identifying which leaves are used for softening the hides and how to locate them, with lower value materials that are used to process the hides, including scraps of the lithic debitage. Around the ages of 8 to 12 years old, the girls are given more responsibility and are allowed to engage in activities such as heat treating the stone used for tool production. It is not until they have reached the ages of 14 to 16 years old that they may begin the process of learning how to make stone tools.

These examples demonstrate how the production of stone tool technology has been treated as a community practice, one that scaffolds the skill acquisition process across many years from early ages. The active learning of the craft does not begin until novices have minimally reached adolescence, when capacities of motor and cognitive ability begin to resemble those of adults. Once this developmental threshold has been met and direct-active teaching has commenced, novice learners continue to hone their craft over at least five years of training, with expert levels of proficiency requiring an even more extensive training period of around a decade (Roux, Bril, and Dietrich 1995). Although experimental studies of stone knapping with modern human participants cannot replicate the processes of ancient hominin learning, these ethnographic case studies suggest that the task of learning to make stone tools was a long one, beginning in childhood with involvement in adjacent-to-knapping subtasks and continuing into adulthood when individuals possessed the anatomical and cognitive affordances to successfully make useful stone tools.

Limitations

As with much of the experimental work conducted on novice stone toolmakers, this study suffers from its relatively short training times and small sample size. An obvious solution to the short training times concern is to conduct an experiment with a longitudinal design, one that tracks cohorts of both children and adults to compare their paths to knapping proficiency. However, given the existing challenges of subject recruitment and retention and raw material costs, one would expect these issues to be magnified in a study design lasting over a period of years (Cotter et al. 2005). It is likely that the resulting sample sizes would be small, and thus, the statistical power of the data would be regrettably weak. These obstacles are what have prevented this sort of work from being done in the past, and to date, the best solutions involve extrapolation of learning trajectories from early-stage novices (Pargeter et al. 2023) and the exploration of substituting comparable, alternative materials (Clarkson 2017; Eren et al. 2022; Khreisheh, Davies, and Bradley 2013; Schillinger, Mesoudi, and Lycett 2014). The ecological validity of such workarounds is tenuous, although one could argue that striving for stringent standards of ecological validity in experimental studies of stone toolmaking is already a moot point, given that this research necessarily uses modern human subjects as surrogates for ancient hominin knappers.

The recruitment and retention of child participants is particularly difficult (Cotter et al. 2002; Prinz et al. 2001), given that a parent or guardian must provide transportation to the experiment site and must either be present at the time of data collection or consent to a drop-off scenario. During the academic year, children enrolled in traditional schooling would be unavailable to engage in research activities during the school day, limiting data collection times to evenings and weekends, which are likely already occupied by extracurricular activities and homework. This restricted availability and its impact on the ability to control knapping practice density is one of the leading factors that motivated this dissertation's summer program design. However, limiting data collection to a single season, while standard practice in archaeological fieldwork, does hinder the rate at which experimental studies, such as this one, can progress.

Although solutions to issues of subject recruitment and retention have been suggested in fields such as developmental psychology (Hurwitz, Schmitt, and Olsen 2017), they may be difficult to implement in experimental archaeological studies due to the public's perceived relative importance of the research questions. In other words, a parent may be more inclined to honor study commitments if they feel that the research benefits either the public, generally, or their child, specifically, as in the case of clinical work. If studies of stone toolmaking in children are viewed as extracurricular, it may be difficult to convince parents and children, alike, that their continued participation is important and a valuable use of their time. To combat this – admittedly cynical – concern, it will be critical provide participants with as much flexibility in scheduling as possible, keeping in mind the importance of practice density in early-stage learning (Pargeter, Khreisheh, and Stout 2019), and to make the experience of learning stone knapping engaging and enjoyable so that children are excited to return for each practice session. To this end, partnerships with organizations that routinely schedule science education outreach programming, such as the Tellus Science Museum, may be one of the most reliable paths forward for experimental studies of stone toolmaking with children.

Conclusion

On Collaborations and Science Education Outreach

One of the most enriching components of my dissertation research was my collaboration with the Tellus Science Museum, in Cartersville, GA. It was no easy task to find a collaborator who would support me in collecting data on how children learned to make stone tools. In my early searching, I was often turned away due to concerns regarding the safety of the task and the liability issues that might follow it. This, in combination with competing priorities of accommodating the needs of my experimental design and maintaining the structure of potential collaborators' existing programming, made finding a compatible collaborator quite difficult, indeed. With the additional challenges of the global pandemic, I feel very fortunate that the Tellus education department staff not only offered the use of their facilities, but they also emphasized from the outset that we were in partnership with one another and made available to me both their material resources (e.g. classroom materials, collections) and their time and expertise. Together, we created a curriculum for what would become Archaeology Adventure Week, and through this experience, I learned how to articulate scientific concepts to audiences of diverse ages.

Not all dissertation projects lend themselves to science education outreach quite as readily as mine does, nor would all dissertation writers be as enthusiastic at the opportunity incorporate elements of public-facing engagement into their work. However, I think there is incredible value not just in establishing ties to the community through collaborations with its educational institutions, but also in challenging oneself to find ways to make the subject of one's research relatable and engaging to general audiences. Through such a symbiosis, it may be possible to create a lasting reciprocal relationship that would essentially become a field site for future studies of stone toolmaking with child participants.

Future Directions

That children performed more poorly than adults at the stone toolmaking task should not be taken as an indication that future studies with developmental populations should be avoided. If anything, it highlights the importance of diverse study designs, especially those that include longer training times or, ideally, a longitudinal timeframe. Because participants in either condition of this experiment did not improve their lithic Quality across the five days of data collection – and, in fact, experienced a reduction in Quality – it is clear that 2.5 hours of training were insufficient for them to move past the novice stage of knapping. It is possible that, given more time, the initial influence of grip strength on lithic Quality would have been mediated by cognitive ability and individuals with higher psychometric scores would have reached local optima similar to that of their stronger counterparts.

Future work on this subject should additionally begin to explore the effects of learning conditions on child knapper outcomes. This study omitted such a design in favor of a proof of concept – it was my priority to demonstrate that children *could*, in fact, produce stone tools reliably enough to compare to an adult sample. It was my fear that introducing learning environments with variable levels of instruction would discourage the children from engaging in the task, altogether, if they, for example, became frustrated by an instructor's lack of responsiveness during a 'no verbal instruction, observation only' condition. Now that it has been shown that children can learn to knap in an experimental setting, it is time to increase the complexity of the experimental design (e.g. modeling Morgan et al. 2015) so that future research can more directly address questions relating to the TOoL hypothesis of language origins.

Final Remarks

Experimental archaeology has grown immensely as a field over recent decades. Like any science, research design and implementation comes with methodological challenges, and in the case of archaeology, these often relate to the deep, evolutionary timescale of the questions asked. Great strides have been made to address these obstacles, but much work still remains to be done. In the case of children's visibility in the archaeological record, it will be critical to refine techniques on identifying individual knappers within lithic assemblages. This is a tall order, and not one I am equipped to engage with here. However, if such a hurtle can be overcome, the possibilities for learning more about not only who made stone tools but also how and when individuals started learning the craft would be immense. This information is vital to gaining a better understanding of how teaching – and, perhaps, by extension – language evolved within our species.

WORKS CITED

Bamforth, Douglas B., and Nyree Finlay. 2008. Introduction: Archaeological Approaches to Lithic Production Skill and Craft Learning. *Journal of Archaeological Method and Theory* 15 (1):1-27.

Baxter, Jane Eva. 2022. *The Archaeology of Childhood*. Second ed: Rowman & Littlefield.

Bird, Douglas W., and Rebecca Bliege Bird. 2000. The Ethnoarchaeology of Juvenile Foragers: Shellfishing Strategies among Meriam Children. *Journal of Anthropological Archaeology* 19 (4):461-476.

Bird, Douglas W.

Bliege Bird, Rebecca. 2002. Children on the Reef: Slow Learning or Strategic Foraging? *Human Nature* 13:269-297.

Bril, B., R. Rein, T. Nonaka, F. Wenban-Smith, and G. Dietrich. 2010. The role of expertise in tool use: skill differences in functional action adaptations to task constraints. *J Exp Psychol Hum Percept Perform* 36 (4):825-39.

Busato, Vittorio V., Frans J. Prins, Jan J. Elshout, and Christiaan Hamaker. 1999. The relation between learning styles, the Big Five personality traits and achievement motivation in higher education. *Personality and Individual Differences* 26:129-140.

Caprara, G. V., M. Vecchione, G. Alessandri, M. Gerbino, and C. Barbaranelli. 2011. The contribution of personality traits and self-efficacy beliefs to academic achievement: a longitudinal study. *Br J Educ Psychol* 81 (Pt 1):78-96.

Cherney, Isabelle D., and Nicole L. Neff. 2004. Role of Strategies and Prior Exposure in Mental Rotation. *Perceptual and Motor Skills* 98:1269-1282.

Clarkson, Chris. 2017. Teaching Complex Flint Knapping Strategies in the Classroom Using "Potato Knapping". *Lithic Technology* 42 (4):155-160.

Costa Jr., Paul T., and McCrae. 1995. Domains and Facets: Hierarchical Personality Assessment Using the Revised NEO Personality Inventory. *Journal of Personality Assessment* 64:21-50.

Cotter, Robert B., Jeffery D. Burke, Rolf Loeber, and Judith L. Navratil. 2002. Innovative Retention Methods in Longitudinal Research: A Case Study of the Developmental Trends Study. *Journal of Child and Family Studies* 11:485-498.

Cotter, Robert B., Jeffrey D. Burke, Magda Stouthamer-Loeber, and Rolf Loeber. 2005. Contacting participants for follow-up: how much effort is required to retain participants in longitudinal studies? *Evaluation and Program Planning* 28 (1):15-21.

Derevensk, J.S. 2000. *Children and Material Culture*: Routledge.

Dreyfus, Stuart E. 2016. The Five-Stage Model of Adult Skill Acquisition. *Bulletin of Science, Technology & Society* 24 (3):177-181.

Eren, Metin I., Lawrence Mukusha, Julie Lierenz, Michael Wilson, Michelle R. Bebber, Michael Fisch, Trent True, Michael Kavaulic, Robert S. Walker, Briggs Buchanan, and Alastair Key. 2022. Another tool in the experimental toolbox: On the use of aluminum as a substitute for chert in North American prehistoric ballistics research and beyond. *North American Archaeologist* 43 (2):151-176.

Estes, Z., and S. Felker. 2012. Confidence mediates the sex difference in mental rotation performance. *Arch Sex Behav* 41 (3):557-70.

Faisal, A., D. Stout, J. Apel, and B. Bradley. 2010. The manipulative complexity of Lower Paleolithic stone toolmaking. *PLoS One* 5 (11):e13718.

Ferguson, Jeffrey R. 2008. The When, Where, and How of Novices in Craft Production. *Journal of Archaeological Method and Theory* 15 (1):51-67.

Fitts, Paul M. 1954. The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement. *Journal of Experimental Psychology* 47:381-391.

Gamble, Clive, and Martin Porr. 2005. From empty spaces to lived lives - Exploring the individual in the Paleolithic. In *Hominid Individual in Context: Archaeological Investigations of Lower and Middle Paleolithic Landscapes, Locales and Artegacts*, edited by C. Gamble and M. Porr: Routledge.

Gensowski, Miriam, Mette Gortz, and Stefanie Schurer. 2021. Inequality in Personality Over the Life Cycle. *CEBI Working Paper Series*.

Geribàs, Núria, Marina Mosquera, and Josep Maria Vergès. 2010. What novice knappers have to learn to become expert stone toolmakers. *Journal of Archaeological Science* 37 (11):2857-2870.

Guida, A., F. Gobet, H. Tardieu, and S. Nicolas. 2012. How chunks, long-term working memory and templates offer a cognitive explanation for neuroimaging data on expertise acquisition: a two-stage framework. *Brain Cogn* 79 (3):221-44.

Hecht, E. E., J. Pargeter, N. Khreisheh, and D. Stout. 2023. Neuroplasticity enables bio-cultural feedback in Paleolithic stone-tool making. *Sci Rep* 13 (1):2877.

Hirtz, Peter, and Wlodzimierz Starosta. 2002. Sensitive and critical periods of motor co-ordination development and its relation to motor learning. *Journal of human kinetics* 7:19-28.

Högberg, Anders. 2018. Approaches to Children's Knapping in Lithic Technology Studies. *Revista de Arqueologia* 31:58-74.

Hurwitz, L. B., K. L. Schmitt, and M. K. Olsen. 2017. Facilitating Development Research: Suggestions for Recruiting and Re-Recruiting Children and Families. *Front Psychol* 8:1525.

Ingold, Tim. 2001. Beyond Art and Technology: The Anthropology of Skill. In *Anthropological Perspectives on Technology*, edited by M. B. Schiffer: University of New Mexico Press.

Kail, Robert, James Pellegrino, and Philip Carter. 1980. Developmental Changes in Mental Rotation. *Journal of Experimental Child Psychology* 29:102-116.

Kerr, Robert. 1975. Movement Control and Maturation in Elementary-Grade Children. *Perceptual and Motor Skills* 41:151-154.

Key, A. J., and C. J. Dunmore. 2015. The evolution of the hominin thumb and the influence exerted by the non-dominant hand during stone tool production. *J Hum Evol* 78:60-9.

Khreisheh, N., D. Davies, and B. Bradley. 2013. Extending Experimental Control: The Use of Porcelain in Flaked Stone Experimentation. *Advances in Archaeologial Practice*:37-46.

Komarraju, Meera, Steven J. Karau, and Ronald R. Schmeck. 2009. Role of the Big Five personality traits in predicting college students' academic motivation and achievement. *Learning and Individual Differences* 19 (1):47-52.

Leone, M. J., D. Fernandez Slezak, G. A. Cecchi, and M. Sigman. 2014. The geometry of expertise. *Front Psychol* 5:47.

Li, Li, Jonathan S. Reeves, Sam C. Lin, Claudio Tennie, and Shannon P. McPherron. 2022. Quantifying knapping actions: a method for measuring the angle of blow on flakes. *Archaeological and Anthropological Sciences* 14 (8).

Lukacs, A., and F. Kemeny. 2015. Development of different forms of skill learning throughout the lifespan. *Cogn Sci* 39 (2):383-404.

Luna, Beatriz, Krista E. Garger, Trinity A. Urban, Nicole A. Lazar, and John A. Sweeney. 2004. Maturation of Cognitive Processes from Late Childhood to Adulthood. *Child Development* 75:1357-1372.

Measelle, J. R., O. P. John, J. C. Ablow, P. A. Cowan, and C. P. Cowan. 2005. Can children provide coherent, stable, and valid self-reports on the big five dimensions? A longitudinal study from ages 5 to 7. *J Pers Soc Psychol* 89 (1):90-106.

Moore, Mark W. 2020. Hominin Stone Flaking and the Emergence of 'Top-down' Design in Human Evolution. *Cambridge Archaeological Journal* 30 (4):647-664.

Morgan, T. J., N. T. Uomini, L. E. Rendell, L. Chouinard-Thuly, S. E. Street, H. M. Lewis, C. P. Cross, C. Evans, R. Kearney, I. de la Torre, A. Whiten, and K. N. Laland. 2015. Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nat Commun* 6:6029.

Nazareth, A., A. Herrera, and S. M. Pruden. 2013. Explaining sex differences in mental rotation: role of spatial activity experience. *Cogn Process* 14 (2):201-4.

Nonaka, T., B. Bril, and R. Rein. 2010. How do stone knappers predict and control the outcome of flaking? Implications for understanding early stone tool technology. *J Hum Evol* 59 (2):155-67.

Nowell, April. 2021. *Growing Up in the Ice Age: Fossil and Archaeological Evidence of the Lived Lives of Plio-Pleistocene Children*: Oxbow.

Pargeter, J., N. Khreisheh, and D. Stout. 2019. Understanding stone tool-making skill acquisition: Experimental methods and evolutionary implications. *J Hum Evol* 133:146-166.

Pargeter, Justin, Cheng Liu, Megan Beney Kilgore, Aditi Majoe, and Dietrich Stout. 2023. Testing the Effect of Learning Conditions and Individual Motor/Cognitive Differences on Knapping Skill Acquisition. *Journal of Archaeological Method and Theory*.

Parsons, T. D., P. Larson, K. Kratz, M. Thiebaux, B. Bluestein, J. G. Buckwalter, and A. A. Rizzo. 2004. Sex differences in mental rotation and spatial rotation in a virtual environment. *Neuropsychologia* 42 (4):555-62.

Prinz, Ronald J., Emilie P. Smith, Jean E. Dumas, James E. Laughlin, Donita W. White, and Rachel Barrón. 2001. Recruitment and Retention of Participants in Prevention Trials Involving Family-Based Interventions. *American Journal of Preventative Medicine* 20:31-37.

Putt, S. S. J., S. Wijeakumar, and J. P. Spencer. 2019. Prefrontal cortex activation supports the emergence of early stone age toolmaking skill. *Neuroimage* 199:57-69.

Reber, Arthur S. 1989. Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General* 118 (3):219-235.

Richardson, John T.E. 1994. Gender Differences in Mental Rotation. *Perceptual and Motor Skills* 78:435-448.

Roux, V., B. Bril, and G. Dietrich. 1995. Skills and learning difficulties involved in stone knapping: The case of stone-bead knapping in Khambhat, India. *World Archaeology* 27:63-87.

Schillinger, Kerstin, Alex Mesoudi, and Stephen J. Lycett. 2014. Copying Error and the Cultural Evolution of "Additive" vs. "Reductive" Material Traditions: An Experimental Assessment. *American Antiquity* 79:128-143.

Schmitt, D. P., A. Realo, M. Voracek, and J. Allik. 2008. Why can't a man be more like a woman? Sex differences in Big Five personality traits across 55 cultures. *J Pers Soc Psychol* 94 (1):168-182.

Scovel, Thomas. 2003. A Critical Review of the Critical Period Research. *Annual Review of Applied Linguistics* 20:213-223.

Snow, Catherine E., and Marian Hoefnagel-Höhle. 1978. The Critical Period for Language Acquisition: Evidence from Second Language Learning. *Child Dev* 49:1114-1128.

Solum, M., H. Loras, and A. V. Pedersen. 2020. A Golden Age for Motor Skill Learning? Learning of an Unfamiliar Motor Task in 10-Year-Olds, Young Adults, and Adults, When Starting From Similar Baselines. *Front Psychol* 11:538.

Soto, C. J., O. P. John, S. D. Gosling, and J. Potter. 2011. Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample. *J Pers Soc Psychol* 100 (2):330-48.

Srivastava, S., O. P. John, S. D. Gosling, and J. Potter. 2003. Development of personality in early and middle adulthood: set like plaster or persistent change? *J Pers Soc Psychol* 84 (5):1041-53.

Stout, D. 2002. Skill and Cognition in Stone Tool Production: An Ethnographic Case Study from Irian Jaya. *Current Anthropology* 43:693-722.

Stout, D., N. Toth, K. Schick, and T. Chaminade. 2008. Neural correlates of Early Stone Age toolmaking: technology, language and cognition in human evolution. *Philos Trans R Soc Lond B Biol Sci* 363 (1499):1939-49.

Stout, Dietrich, Nicholas Toth, Kathy Schick, Julie Stout, and Gary Hutchins. 2000. Stone Tool-Making and Brain Activation: Position Emission Tomography (PET) Studies. *Journal of Archaeological Science* 27 (12):1215-1223.

Stroop, J. Ridley. 1935. Studies of Interference in Serial Verbal Reactions. *Journal of Experimental Psychology* 18:643-662.

Toth, Nicholas, and Kathy Schick. 2018. An overview of the cognitive implications of the Oldowan Industrial Complex. *Azania: Archaeological Research in Africa* 53 (1):3-39.

Vandenberg, Steven G., and Allan R. Kuse. 1978. Mental Rotations, A Group Test of Three-Dimensional Spatial Visualization. *Perceptual and Motor Skills* 47:599-604.

Vernooij, Carlijn A., Leonora J. Mouton, and Raoul M. Bongers. 2012. Learning to Control Orientation and Force in a Hammering Task. *Zeitschrift für Psychologie* 220 (1):29-36.

Weedman Arthur, Kathryn. 2010. Feminine Knowledge and Skill Reconsidered: Women and Flaked Stone Tools. *American Anthropologist* 112 (2):228-243.

Williams-Hatala, E. M., K. G. Hatala, M. Gordon, A. Key, M. Kasper, and T. L. Kivell. 2018. The manual pressures of stone tool behaviors and their implications for the evolution of the human hand. *J Hum Evol* 119:14-26.

Williams-Hatala, E. M., K. G. Hatala, A. Key, C. J. Dunmore, M. Kasper, M. Gordon, and T. L. Kivell. 2021. Kinetics of stone tool production among novice and expert tool makers. *Am J Phys Anthropol* 174 (4):714-727.

Williams, B.R., J.S. Ponesse, R.J. Schachar, G.D. Logan, and R. Tannock. 1999. Development of Inhibitory Control Across the Life Span. *Developmental Psychology* 35:205-213.

Wynn, T. 2002. Archaeology and cognitive evolution. *Behavioral and Brain Sciences* 25:389-438.

Wynn, T., and F. L. Coolidge. 2016. Archeological insights into hominin cognitive evolution. *Evol Anthropol* 25 (4):200-13.

Wynn, T., R. A. Hernandez-Aguilar, L. F. Marchant, and W. C. McGrew. 2011. "An ape's view of the Oldowan" revisited. *Evol Anthropol* 20 (5):181-97.