How we value the assistance

on each Cowork-attributed task

Per-category bands from independent research. Summary table with linked sources. Per-category derivations, one slide per bucket.

THE LEARNING JOURNEY

Assessing the research methodology

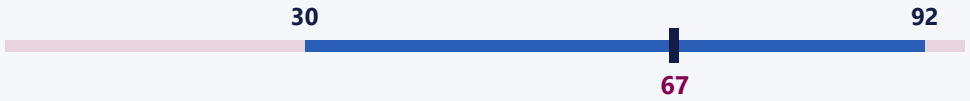
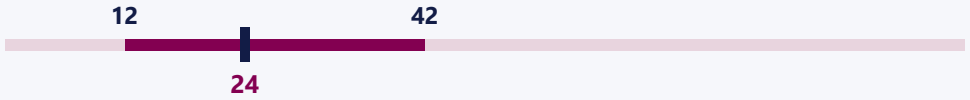
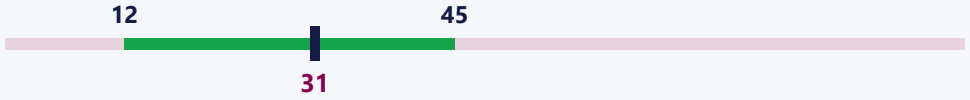
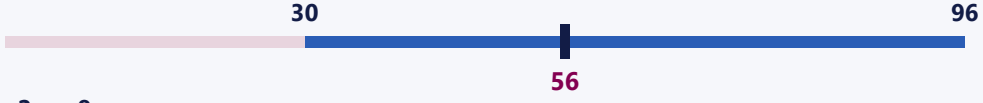
Each task category gets a minutes-saved range, traceable to a published study.

WHAT YOU'LL BE ABLE TO ANSWER AFTER THIS DECK

- **Q1** What are the categories of tasks?
- **Q2** How is each task category's band derived — and from which study?
- **Q3** How do we envision these with a scenario ?

Each task category is independently assessed — H / M / L bands

Bands come from independent published research per category. Not vendor self-report. Not extrapolation.

TASK CATEGORY	BAND (min/task) ◀ LOW —— TYPICAL —— HIGH ▶	WHY THIS 'TYPICAL' · derivation (study × instances per task)
1 Analysis & Research		*LOW = McKinsey 2023 (25–40% uplift × ≈60-min baseline analytical task ≈ 30 min, constructed). MID = Stanford-WB 2025 mean of the 5 research-adjacent O*NET categories (Critical Thinking · Active Learning · Quality Control Analysis · Judgement & Decision Making · Complex Problem Solving) = 335/5 ≈ 67 min/task. HIGH = highest single category (Complex Problem Solving) = 92 min/task.*
2 Document & content creation		*Microsoft Research DiD 2026 (TIER-1): 6.1 min/Word activity instance (n=72,186) × 4 instances per doc run (draft → rewrite → format → polish) ≈ 24*
3 Email workflows		*Dillon et al. NBER w33795 (2025): ~2 hr/wk email-time savings (abstract) ÷ 14.5 replies/wk (Table 2 pre-period) ≈ 7–8 min/reply.*
4 Meeting workflows		*Microsoft WTI Special Report 2023 (Study #2, Cambon et al. methodology): n=57 RCT, recap of a 35-min missed Teams meeting. Without Copilot 42m 34s → with Copilot 11m 13s ≈ 31 min saved per recap (~3.8× faster).*
5 Communication workflows		*WTI 2024: 2 min/comms instance × 2 instances per comms run (rule + AI draft) ≈ 4*
6 Specialized workflows		*Forrester TEI Power Automate 2024: ~32 min/full automation; trimmed to 25 because typical Cowork specialized run has fewer cross-system steps than full Power Automate workflow (conservative)*
7 Write or debug code		*Cui CACM 2024 (n=4,867): ~18 min saved per coding step (26% × ~70 min task) × 3 steps per code run (write + test + debug) ≈ 56 (not the 30↔96 midpoint of 63)*
8 General assistance / Other		*WTI 2024 + Brynjolfsson QJE 2025: ~5 min per assist/learn episode (single instance, no chain — this bucket is by definition non-chained)*

Note: Each "Mid/Typical" value is calculated as the sum of per-instance metrics for the specific activities that make up a task, based on a given study. The right column shows the calculation behind each "Mid."

Analysis & Research

BAND (min/task)

30 → 67 → 92

DERIVATION · exactly how this Mid was computed

MID COMPUTATION

MID (67) = Stanford-WB 2025: mean of the 5 O*NET categories labeled on the chain below — CT 75 · AL 50 · QCA 67 · J&DM 51 · CPS 92 → 335 ÷ 5 ≈ 67 min/task. HIGH (92) = CPS (highest single category). LOW (30) = McKinsey 2023 (see right anchor).

SOURCE EXCERPT · key finding paraphrased

*Stanford-WB 2025 (Hartley · Jolevski · Melo · Moore, n=4,278): per-task savings across 18 O*NET categories. The 5 analysis-research categories average 67 min/task (range 50–92). McKinsey 2023: 25–40% uplift on analytical work → ~30 min saved on a 60-min baseline (LOW anchor). No RCT publishes a per-analysis-task minute floor at this scale today — 30 used as conservative floor.*

Sources → Stanford-WB SSRN 5136877 (2025) ↗ · McKinsey 2023 ↗

TYPICAL ACTIVITY-INSTANCE CHAIN · cognitive phases of an analysis run, each labeled with its Stanford-WB O*NET category

WORKED EXAMPLE · what an agentic Cowork run looks like in this category

"Pull last quarter's churn data, identify the top drivers, cross-check against NPS verbatims, and give me a one-pager for Friday's QBR."

WITHOUT COWORK ≈ 85 min

Export from CRM · build pivots · search NPS verbatims for themes · hand-write up findings

WITH COWORK ≈ 18 min

Describe scope · review the synthesis the agent assembled · tighten framing

= ≈ 67 min saved per run (matches Typical band of 67 min)

RESEARCH ANCHORS · one per band point

LOW · 30 min

McKinsey 2023 (derived)

Construction: 25–40% McKinsey uplift × ~60-min baseline analytical task ≈ 30 min. No RCT measures a per-analysis-task minute floor today — transparent estimate.

MID · 67 min

Stanford-WB 2025 (basket mean)

*Each chain step is labeled with one Stanford-WB O*NET category.* ***67 = mean of the basket (CT 75 · AL 50 · QCA 67 · J&DM 51 · CPS 92) — not a sum across steps.***

HIGH · 92 min

Stanford-WB 2025 (highest)

Highest of the 5 research-adjacent categories — cross-domain analysis ceiling.

Document & content creation

BAND (min/task)

12 → **24** → 42

★ TIER-1 ANCHOR — Microsoft Research Research causal-impact (DiD, future-adopter control, n=72,186 users, ~6.1 min saved per Word activity instance).

DERIVATION · exactly how this Mid was computed

MID COMPUTATION

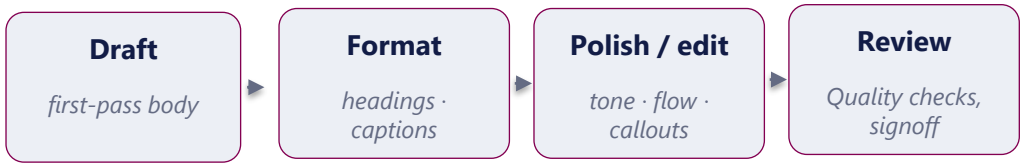
Microsoft Research DiD 2026 (TIER-1): 6.1 min/Word activity instance (n=72,186) × 4 instances per doc run (draft → rewrite → format → polish) ≈ 24

SOURCE EXCERPT · key finding paraphrased

Verma, Suri & Counts (2026), Microsoft Research causal-impact study: across 72,186 Word users in 1,021 US tenants, with a difference-in-differences design using future adopters as control, Copilot users saved ~6.1 minutes per Word activity instance (p<.001). Real doc runs are not one-shot — they typically chain ~4 Word-activity instances (initial draft, targeted rewrites, formatting pass, polish). UK GDS 2025 (n=20K civil servants) independently reports ~24 min saved per drafting task, validating the Mid.

Sources → Microsoft Research Research 2026 (Verma · Suri · Counts) ↗

TYPICAL ACTIVITY-INSTANCE CHAIN · intra-bucket



WORKED EXAMPLE · what an agentic Cowork run looks like in this category

"Turn the churn analysis above into a polished exec one-pager — clean tone, chart captions, and a callout box for the recommendation."

WITHOUT COWORK ≈ 30 min

Draft and rewrite weak sections · format headings + captions · polish language · sanity-check flow

WITH COWORK ≈ 6 min

Review the draft · ask for one section rewrite · light edits in place

= **≈ 24 min saved per run** *(matches Typical band of 24 min)*

RESEARCH ANCHORS · one per band point

LOW · 12 min

Microsoft Research × 2 instances

Short doc: quick draft + light edit

MID · 24 min

Microsoft Research × 4 instances

Draft → rewrite → format → polish; corroborated by UK GDS ~24 min/drafting task (n=20K)

HIGH · 42 min

Microsoft Research × 7 instances

Long-form doc with multiple rewrite cycles + heavy in-place editing

Email workflows

BAND (min/task)

3 → 7 → 12

DERIVATION · exactly how this Mid was computed

MID COMPUTATION

Dillon et al. NBER w33795 (2025): ~2 hr/wk email-time savings (abstract) ÷ 14.5 replies/wk (Table 2 pre-period) ≈ 7–8 min/reply. Used directly as Typical (7 min, conservative).

SOURCE EXCERPT · key finding paraphrased

Dillon, Jaffe, Peng & Cambon (2025), NBER Working Paper w33795: randomized field experiment, n=6,000+ workers across 56 firms, 6-month deployment of M365 Copilot. The abstract reports Copilot-treated workers saved approximately 2 hours per week on email. Table 2 pre-period mean = 14.5 unique threads replied to per week. 2 hr/wk ÷ 14.5 replies/wk ≈ 7–8 min per substantive reply. Noy & Zhang 2023 (Science 381, 187) provides the lower bound: a single quick AI-drafted reply ≈ 3 min.

Sources → Dillon et al. NBER w33795 (2025) ↗ · Noy & Zhang, Science 2023 ↗

TYPICAL ACTIVITY-INSTANCE CHAIN · intra-bucket

Read prior threads — *Read prior thread* ▸ **Gather additional context** — *Synthesis* ▸ **Respond** — *AI draft accept*

WORKED EXAMPLE · what an agentic Cowork run looks like in this category

"Catch me up on the contract renewal thread (8 messages, 2 attachments) and draft a reply that holds our pricing — flag any commitment risks against the SOW."

WITHOUT COWORK ≈ 10 min

Read all 8 messages · open both attachments · cross-check the SOW · draft carefully

WITH COWORK ≈ 3 min

Review the thread summary, the draft, and the commitment-risk flags it surfaced

= ≈ 7 min saved per run (matches Typical band of 7 min)

RESEARCH ANCHORS · one per band point

LOW · 3 min

Noy & Zhang Science 2023 RCT

~40% reduction (~11 min faster, 27→17 min) on writing → ~3 min/email scaled

MID · 7 min

Dillon NBER w33795 (2025)

2 hr/wk savings (abstract) ÷ 14.5 replies/wk (Table 2) ≈ 7–8 min/reply.

HIGH · 12 min

Dillon NBER w33795 (2025)

Heavy-thread reply chain (multi-message context + attachment reasoning) for top-decile email volume users.

Meeting workflows

BAND (min/task)

12 → 31 → 45

DERIVATION · exactly how this Mid was computed

MID COMPUTATION

Microsoft WTI Special Report 2023 (Study #2; Cambon et al. methodology): n=57 RCT, recap of a 35-min missed Teams meeting. Without Copilot 42m 34s → with Copilot 11m 13s = ≈ 31 min saved per recap (~3.8× faster). Used directly as Typical.

SOURCE EXCERPT · key finding paraphrased

Microsoft Work Trend Index Special Report (Nov 2023, 'What Can Copilot's Earliest Users Teach Us About Generative AI at Work?') reports Study #2: 57 Microsoft employees split into Copilot and control groups, each asked to summarize a 'missed' 35-minute Teams meeting (with crosstalk and debate to mirror real meetings). Copilot users completed the recap in 11m 13s vs. 42m 34s for control — ~31 min saved per recap, 3.8× faster, 58% less draining. Cambon et al. 2023 (Microsoft Research) is the companion methodology paper for the same experiment. Caveat: the saving is for the act of catching up on a missed meeting via transcript+recording — a task many people may not perform regularly without AI.

Sources → Microsoft WTI Special Report 2023 ↗ · Cambon et al., Microsoft Research 2023 ↗

TYPICAL ACTIVITY-INSTANCE CHAIN · intra-bucket

Multi-meeting pull	→	Theme synthesis	→	Drift detection	→	Targeted actions
last 3 syncs on topic		*decisions · blockers · open Qs*		*what changed across meetings*		*2–3 cross-tool follow-ups*

WORKED EXAMPLE · what an agentic Cowork run looks like in this category

"We've had 3 design syncs on the auth rewrite over the last 2 weeks. Pull them all, surface what we actually decided vs. what's still open, flag where we contradicted ourselves, and post a consolidated update to the team channel with the 2 open questions tagged to owners."

WITHOUT COWORK ≈ 37 min

Re-scrub 3 recordings · reconcile decisions across them · spot contradictions manually · hunt down owners · draft + post the update

WITH COWORK ≈ 6 min

Review the cross-meeting synthesis, the contradiction flags, and the draft channel post — approve owners and send

= ≈ 31 min saved per run (matches Typical band of 31 min)

RESEARCH ANCHORS · one per band point

LOW · 12 min

Cambon MSR 2023 (single recap)

Short / well-structured meeting recap or single action-item extraction — conservative lower bound.

MID · 31 min

Microsoft WTI 2023 · Study #2

n=57 RCT recapping a 35-min missed Teams meeting: 42m 34s without Copilot → 11m 13s with Copilot ≈ 31 min saved.

HIGH · 45 min

Cross-meeting agentic chain

31 (WTI #2 RCT, one missed-meeting recap, measured) + 14 (Cambon MSR 2024, daily meeting-related AI assistance per user, measured). Both Microsoft-measured. No extrapolation

Communication workflows

BAND (min/task)

2 → 4 → 6

DERIVATION · exactly how this Mid was computed

MID COMPUTATION

WTI 2024: 2 min/comms instance × 2-3 instances per comms run (rule + AI draft) ≈ 4

SOURCE EXCERPT · key finding paraphrased

Microsoft Work Trend Index 2024: ~14 min/day saved on communication tasks (Teams, channel routing, status replies). Per-micro-task savings ≈ 2 min derived as 14 min/day ÷ 5–7 micro-tasks/day.

Sources → Microsoft WTI 2024 ↗ · Dillon et al. NBER w33795 (2025) ↗

TYPICAL ACTIVITY-INSTANCE CHAIN · intra-bucket

Teams reply	Channel rule	Multi-channel send
AI suggested reply	*AI-drafted rule*	*rule + draft + send*

WORKED EXAMPLE · what an agentic Cowork run looks like in this category

"I'm OOO Friday — scan my channel mentions this week, draft handoff notes in each open thread, and queue an OOO message in my standup channel."

WITHOUT COWORK ≈ 6 min

Scroll mentions across 4 channels · write a handoff per thread · set OOO manually

WITH COWORK ≈ 2 min

Review the drafts in place · post / approve

= **≈ 4 min saved per run** *(matches Typical band of 4 min)*

RESEARCH ANCHORS · one per band point

LOW · 2 min

Microsoft WTI 2024

14 min/day comms ÷ 5–7 micro-tasks/day = 2 min/instance

MID · 4 min

Microsoft WTI 2024

× 2 instances (rule + draft + send)

HIGH · 6 min

NBER

216 min/wk comms savings ÷ ~35 project threads as a proxy for channels and group chats (7 "working spheres"/day × 5) ≈ 6 min each. Sources: Mark, Attention Span 2023; corroborated by Dillon et al., NBER w33795 (2025).

Specialized workflows

BAND (min/task)

10 → 25 → 40

DERIVATION · exactly how this Mid was computed

MID COMPUTATION

Forrester TEI Power Automate 2024: ~32 min/full automation; trimmed to 25 because typical Cowork specialized run has fewer cross-system steps than full Power Automate workflow (conservative)

SOURCE EXCERPT · key finding paraphrased

Forrester Consulting Total Economic Impact™ of Microsoft Power Automate (2024): composite organization saved 200 hours per knowledge worker per year via cross-system workflow automation. Derivation: 200 hr/yr ÷ 250 working days ÷ ~1.5 workflows/day ≈ 32 min/automation. UK GDS 2025 cross-government (n=20K civil servants) corroborates single-system admin task savings ~10 min.

Sources → Forrester TEI Power Automate 2024 ↗ · UK GDS Cross-Government 2025 ↗

TYPICAL ACTIVITY-INSTANCE CHAIN · intra-bucket

Single system action	Cross-system chain 1	Cross-system chain 2	Confirm + audit
schedule, file, route	*trigger downstream*	*merge → update*	*verify completion*

WORKED EXAMPLE · what an agentic Cowork run looks like in this category

"A new Tier-1 support ticket came in — classify it, pull the customer's contract tier from CRM, draft a reply with the right SLA language, ping the on-call engineer in Teams, and log the action in SharePoint."

WITHOUT COWORK ≈ 30 min

Read ticket · open CRM → look up tier · grab template → draft · switch to Teams · fill SharePoint log row

WITH COWORK ≈ 5 min

Describe intent · review the orchestrated draft + log entry · approve

= **≈ 25 min saved per run** *(matches Typical band of 25 min)*

RESEARCH ANCHORS · one per band point

LOW · 10 min

UK GDS June 2025

Scheduling/admin single-system turn ~10 min

MID · 25 min

Forrester TEI Power Automate 2024

200 hr/yr ÷ 250 days ÷ 1.5 workflows/day ≈ 32 min/automation

HIGH · 40 min

Stanford & World Bank 2025

Time Management 48 min/task; trimmed to 40 as multi-instance ceiling

Write or debug code

BAND (min/task)

30 → 56 → 96

DERIVATION · exactly how this Mid was computed

MID COMPUTATION

Cui CACM 2024 (n=4,867): ~18 min saved per coding step (26% × ~70 min task) × 3 steps per code run (write + test + debug) ≈ 56 (not the 30↔96 midpoint of 63)

SOURCE EXCERPT · key finding paraphrased

Cui, Demirer, Jaffe et al. (2024), Communications of the ACM: across n=4,867 GitHub Copilot users in 3 enterprise field experiments, completed pull requests rose 26.08% (95% CI: 21.4–30.7%). Time saved per coding step ≈ 18 min on tasks averaging ~70 min. Peng et al. 2023 RCT on n=95 developers: 55.8% time reduction on a controlled coding task (71 min → 31 min). Stanford-WB 2025 sets the ceiling.

Sources → Cui et al., CACM 2024 ↗ · Peng et al. RCT 2023 (arXiv) ↗ · Stanford-WB SSRN 5136877 ↗

TYPICAL ACTIVITY-INSTANCE CHAIN · intra-bucket

Write code	Tests	Debug	Commit	Review
feature or fix	*draft + run*	*stack trace + fix*	*stage + push*	*PR comments / iterate*

WORKED EXAMPLE · what an agentic Cowork run looks like in this category

"Build a Python script that predicts churn from the CRM export — train/test split, basic feature engineering, and a short eval report."

WITHOUT COWORK ≈ 70 min

Scaffold the script · write feature code · train · debug edge cases · hand-write the eval report

WITH COWORK ≈ 14 min

Review the generated script · run · iterate on edge cases the agent flagged

= **≈ 56 min saved per run** *(matches Typical band of 56 min)*

RESEARCH ANCHORS · one per band point

LOW · 30 min

Peng et al. 2023 GitHub Copilot RCT

55.8% reduction on 71 min task ≈ 31 min × 2 instances

MID · 56 min

Cui et al. CACM 2024 (n=4,867)

26% productivity gain × ~3 instances of 18 min

HIGH · 96 min

Stanford & World Bank 2025

96 min/task ceiling; 5 instances × 20 min/instance

General assistance / Other

BAND (min/task)

2 → 5 → 8

DERIVATION · exactly how this Mid was computed

MID COMPUTATION

WTI 2024 + Brynjolfsson QJE 2025: ~5 min per assist/learn episode (single instance, no chain — this bucket is by definition non-chained)

SOURCE EXCERPT · key finding paraphrased

Brynjolfsson, Li & Raymond (2025), Quarterly Journal of Economics / NBER w31161: AI assistance raised customer-service agent productivity by 14% on average and 34% for novice agents (n=5,179 agents, 5M+ conversations). Translates to a typical single-turn assist/learn episode of ~5 min saved.

Sources → Brynjolfsson, Li & Raymond QJE 2025 / NBER w31161 ↗ · Microsoft WTI 2024 ↗

TYPICAL ACTIVITY-INSTANCE CHAIN · intra-bucket

Single Q&A — *quick help* ▸ **Learning episode** — *explain / teach* ▸ **Micro-task** — *format / convert / lookup*

WORKED EXAMPLE · what an agentic Cowork run looks like in this category

"Reformat this messy list of ~80 customer names into a clean comma-separated string — deduplicated and sorted alphabetically."

WITHOUT COWORK ≈ 7 min
Paste into editor · manually dedupe · sort · join with commas

WITH COWORK ≈ 2 min
Paste the list · describe the format · verify the output

= **≈ 5 min saved per run** *(matches Typical band of 5 min)*

RESEARCH ANCHORS · one per band point

LOW · 2 min

Microsoft WTI 2024

Micro-Q&A floor for single-turn micro-tasks

MID · 5 min

WTI + Brynjolfsson QJE 2025

Typical assist/learn episode

HIGH · 8 min

Brynjolfsson, Li & Raymond QJE / NBER w31161

Novice-learning episode (14% productivity gain in low-tenure)

SUMMARY TABLE

All 8 categories at a glance — Low / Typical / High with key sources

TASK CATEGORY	LOW	TYPICAL	HIGH	KEY SOURCE(S) FOR THE TYPICAL VALUE
1 Analysis & Research *~67 min / analysis run (Stanford-WB basket mean)*	30	67	92	Stanford-WB SSRN 5136877 (2025) ↗ · McKinsey 2023 ↗ *Stanford-WB 2025 (SSRN 5136877): mean of 5 research-adjacent O*NET categories (Critical Thinking 75, Active Learning 50, Quality Control Analysis 67, Judgement & Decision Making 51, Complex Problem Solving 92) = 335/5 ≈ 67 min/task. LOW = McKinsey 2023 construction*
2 Document & content creation *~6.1 min / Word activity instance*	12	24	42	**Microsoft Research Research 2026 (Verma · Suri · Counts) ↗** *Microsoft Research DiD 2026 (TIER-1): 6.1 min/Word activity instance (n=72,186) × 4 instances per doc run (draft → rewrite → format → polish) ≈ 24*
3 Email workflows *~7 min / email-reply instance*	3	7	12	Dillon et al. NBER w33795 (2025) ↗ · Noy & Zhang, Science 2023 ↗ *Dillon NBER w33795 (2025) RCT, n=6,000+ across 56 firms: ~2 hr/wk email-time savings (abstract) ÷ 14.5 replies/wk (Table 2 pre-period) ≈ 7–8 min/reply. Used directly as Typical (7, conservative).*
4 Meeting workflows *~31 min / missed-meeting recap*	12	31	45	Microsoft WTI Special Report 2023 ↗ · Cambon et al., MSR 2023 ↗ *Microsoft WTI Special Report 2023, Study #2 (n=57 RCT, Cambon et al. methodology): recap of a 35-min missed Teams meeting — 42m 34s without Copilot → 11m 13s with Copilot ≈ 31 min saved per recap (~3.8× faster). Used directly as Typical.*
5 Communication workflows *~2 min / non-email comms instance*	2	4	6	Microsoft WTI 2024 ↗ · Dillon et al., NBER w33795 ↗ *WTI 2024: ~14 min/day comms time savings ÷ 5–7 micro-tasks/day ≈ 2 min/instance × ~2 instances (rule + AI draft) ≈ 4. HIGH (6) = 216 min/wk comms savings ÷ ~35 workstreams/wk (7 working spheres/day × 5; Mark, Attention Span 2023; corroborated by Dillon et al., NBER w33795).*
6 Specialized workflows *~10 min / workflow-step instance*	10	25	40	**Forrester TEI Power Automate 2024 ↗** · **UK GDS Cross-Government 2025 ↗** *Forrester TEI Power Automate 2024: ~32 min/full automation; trimmed to 25 because typical Cowork specialized run has fewer cross-system steps than full Power Automate workflow (conservative)*
7 Write or debug code *~15–20 min / coding instance*	30	56	96	**Cui et al., CACM 2024 ↗** · **Peng et al. RCT 2023 (arXiv) ↗** · **Stanford-WB SSRN 5136877 ↗** *Cui CACM 2024 (n=4,867): ~18 min saved per coding step (26% × ~70 min task) × 3 steps per code run (write + test + debug) ≈ 56 (not the 30↔96 midpoint of 63)*
8 General assistance / Other *2–8 min / single-instance assist*	2	5	8	**Brynjolfsson, Li & Raymond QJE 2025 / NBER w31161 ↗** · **Microsoft WTI 2024 ↗** *WTI 2024 + Brynjolfsson QJE 2025: ~5 min per assist/learn episode (single instance, no chain — this bucket is by definition non-chained)*

Note: Each "Mid/Typical" value is calculated as the sum of per-instance metrics for the specific activities that make up a task, based on a given study. The right column shows the calculation behind each "Mid."