Process Improvement and System Design Portfolio

Five Lean Six Sigma projects across education, senior care, assessment and supplier management

Charles A. Hollenbeck, Ph.D., Ed.D.

Lean Six Sigma Master Black Belt · PRINCE2 Practitioner · PSM III · PSPO III · ProKanban II · GRCP · GRCA

Work sample, 2026

What this document is

Five Lean Six Sigma projects, summarized. Each is documented in full at Master Black Belt scope and runs 18 to 32 pages; this is the overview.

Organizations and individuals are described by sector and scale rather than by name. Full documents are available on request where confidentiality permits.

One project is a method demonstration built on constructed data and is labeled as such. The other four use actual operating records.

At a glance

#	PROJECT	FRAMEWORK	SECTOR	SCALE
1	**Asynchronous writing feedback service**	DMAIC	Higher education	8 campuses, 16 programs, distributed team of 8 to 12
2	**Contracted tutoring supplier quality**	DMAIC	Vendor management	1,292 sessions, 514 students, 652 tutors
3	**Resident activity programming capacity**	DMAIC	Senior care, regulated	706 sessions across 7 months, 70-resident population
4	**Constructed-response scoring reliability**	DMAIC	Standardized assessment	Method demonstration, constructed data
5	**Multi-subject instructional content system**	DMADV	Higher education	240+ assets, 6 subjects, 16-person cross-functional team

Methods demonstrated across the portfolio: capability analysis on non-normal and mixture distributions · attribute agreement and measurement system analysis · logistic regression with cluster-robust standard errors · 2^3 factorial design of experiments · Gaussian mixture modeling · statistical process control matched to available detection power · quality function deployment · FMEA with published rating scales · concept selection against datum · bootstrap confidence intervals · power and sample-size analysis.

1. Asynchronous writing feedback service

Higher education · DMAIC · 8 campuses, 16 degree programs

A multi-campus university had no scalable way to give students written feedback on drafts. I designed and launched an asynchronous drop-off service across eight campuses and sixteen degree programs, staged campus by campus, running to a 24 to 48 hour turnaround standard with a distributed team of eight to twelve instructors.

What I found. The service worked. What it could not do was prove it. The intake instrument captured a single timestamp, so the turnaround commitment the whole service was built around was never measurable from the operating record. That became the first instance of a pattern I would find in every subsequent project.

What it demonstrates. Service design from scratch, multi-site rollout, distributed team management, and the discipline to report a measurement gap rather than assert a result the data could not support.

2. Contracted tutoring supplier quality

Vendor management · DMAIC · 1,292 sessions analyzed

An institution had contracted a third-party tutoring provider and had no way to assess whether it was delivering. I built the analysis from the vendor's own session records.

What I found. One live session in five ended within five minutes, a rate stable across nine months and therefore systemic rather than incidental. Failure concentrated in the institution's own core subjects rather than in general education ones, meaning the supplier performed adequately at what any institution teaches and poorly at what this one existed to teach.

Three hypotheses died in the process. Tutor discontinuity looked causal until I traced the apparent effect to reconnection behavior and it collapsed.

Tutor inexperience ran in the opposite direction to prediction and then vanished under adjustment. A modality confound was inflating the apparent performance until stratification removed it, at which point the defect rate rose rather than fell.

What it demonstrates. Supplier quality management under real authority limits, since the institution employed none of the tutors. Measurement system analysis applied to a supplier's own detection system. Willingness to report the number that got worse under scrutiny.

3. Resident activity programming capacity

Senior care, state-regulated · DMAIC · Before-and-after on operating records

A residence serving seventy assisted-living residents ran a limited activity program carried largely by one coordinator. I joined, spent a week observing before changing anything, surveyed residents, held informal conversations about what they actually wanted, and then rebuilt the schedule.

What I found. Programming rose 56%, from 2.26 to 3.51 sessions per day, and held for six months. The increase was additive rather than a reallocation: existing programming continued at 0.91 times its prior rate while sixty-seven new formats were layered on.

Where the capacity came from is the interesting part. Not longer days or more staff. The early afternoon, which carried most of the original program, was unchanged. Early mornings rose sevenfold, mid-afternoons more than tripled, weekends more than doubled. The building had the hours all along and nobody had scheduled into them. One format that existed already and ran once a month became a five-times-a-week staple, requiring no new materials, room or skill.

What I could not establish. Whether the program reached more residents or gave the same residents more to do. The attendance record attributes counts to a calendar cell rather than to a session, so ninety-one percent of sessions cannot be traced to a participant figure. I reported that as the primary constraint on the project rather than working around it.

What it demonstrates. Capacity analysis, before-and-after design against a documented intervention point, control charting matched honestly to available detection power, and a measurement redesign proposed because the existing one could not answer the question.

4. Constructed-response scoring reliability

Standardized assessment · DMAIC · Method demonstration, constructed data

> This project uses simulated data throughout. It is grounded in my experience as a constructed-response scoring professional, and it is a demonstration of method rather than a report of work performed for any employer. The full document discloses this in its opening paragraph.

Human raters applying an ordinal rubric produce variation. Where two raters disagree by two points or more, adjudication follows, score release slows, and a student's classification depends partly on who received their response.

The design. A generative model with documented parameters and seeds produces 1,000 double-scored responses and a repeated-anchor study of 30 anchors across 10 raters, two trials, two phases. Every figure traces to those datasets.

What it demonstrates. Ordinal attribute agreement analysis, correctly chosen over variable gauge R&R because a rubric produces ordered categories rather than equal-interval measurements. Quadratic-weighted kappa with bootstrap intervals. A 2^3 factorial experiment that produced a non-obvious result: an expanded exemplar library achieves almost nothing on its own and only works when combined with a mechanism that makes raters revisit it.

And one target was missed by two tenths of a point. The document records the decision taken in response, declines to lower the threshold to meet it, and carries it into the control phase as an open item. That is deliberate. A pilot in which every measure hits its target is a pilot nobody should believe.

5. Multi-subject instructional content system

Higher education · DMADV · 240+ published assets · 16-person team

An academic resource center had no instructional content: no repository, no production process, no way for student demand to reach anyone able to answer it. Subject expertise, information management capability and a publishing platform all existed in isolation, connected by nothing.

What I built. Over eighteen months, leading a cross-functional team of librarians, subject faculty, information management staff and academic leadership: 83 written guides, 21 numbered explainer videos, and at least 136 faculty-produced subject lectures across six disciplines. Topic selection ran from a monitored stream of student search inquiries. Production was distributed to six subject experts who delivered 1.31 times my own output while carrying full teaching loads, coordinated on a bi-weekly reporting cycle.

The mechanism worth naming. I could not have specified what a microbiology or pathophysiology video should contain. I could specify when it was due, what it should look like, and how its topic would be chosen. Separating the what from the when and how is what allowed one coordinator to direct production across six disciplines he was not qualified to teach.

What I found on review. Eighty-seven percent of the written collection covered topics with freely available published equivalents, representing roughly 300 hours spent reproducing material that already existed. The build-versus-curate decision had been taken once as a policy rather than per

asset against a rule. In a later role facing the same problem I curated instead, at zero production cost, which makes the comparison one I have run rather than one I have argued.

And the system had no feedback. No usage data existed for any asset. Search inquiries flowed in and shaped what was built; nothing flowed back to indicate whether it was found or useful. A design that cannot observe its own performance can produce well, and did, but cannot improve. It can only accumulate.

What it demonstrates. Design for Six Sigma on a capability that did not previously exist. Quality function deployment, concept selection against a datum, design FMEA, verification experiments specified with power analysis. Cross-functional leadership across roles I had no authority over. And a self-assessment that reports the least flattering finding in the project.

Verifiable outcome. The collection remains live and institutionally maintained twenty-five months after my departure.

The pattern across all five

Four organizations, four sectors, five projects. In every one, the organization was recording something faithfully and could not answer the question that mattered.

A writing service that never captured its own turnaround. A supplier report that counted volume and was silent on quality. An activity log that attributed attendance to a calendar cell rather than a session. A content library with no usage data at all. A scoring operation whose exception flag had never been validated.

None of these were failures of diligence. In each case someone was doing the work carefully. What was missing was an instrument capable of seeing whether the work achieved what it was for.

I did not set out to find that. It is what kept turning up, and it is now how I approach any process: establish what the measurement can and cannot see before drawing conclusions from it, and treat a measurement gap as a finding rather than an inconvenience.

Full project documents, methodology appendices and datasets are available on request where confidentiality permits.