Interdisciplinary Journal of Environmental and Science Education
2026, 22(4), e2624
e-ISSN: 2633-6537
https://www.ijese.com/

MODESTUM

Research Article

OPEN ACCESS

Remembering competence: Retrospective self-assessment of sustainability learning four years after a higher education course

André Dutra [1*]

[1] Independent researcher, Uppsala, SWEDEN
*Corresponding Author: andretnies@gmail.com

Citation: Dutra, A. (2026). Remembering competence: Retrospective self-assessment of sustainability learning four years after a higher education course. *Interdisciplinary Journal of Environmental and Science Education, 22*(4), Article e2624. https://doi.org/10.29333/ijese/19770

ARTICLE INFO

Received: 28 Jun. 2026
Accepted: 26 Oct. 2026

ABSTRACT

This article examines how participants in a discontinued sustainability course at a Swedish university characterized their learning four years after it ended, and what remembering itself reveals about the limits of retrospective evaluation in sustainability education research. Drawing on a single-case study, it revisits ten semi-structured interviews with five students, three instructors and two staff members reflecting on a course that had ended four years earlier. Rather than treating recall difficulties as noise, the article repositions them as a central finding. Participants struggled to separate this course from similar courses taken concurrently, and their competency ratings diverged sharply from one another and from the original synthesized account. Contrasted against contemporaneous course evaluation data, the interviews show that retrospective accounts reorganize, conflate and disagree rather than simply fading. The article argues that single-case retrospective designs should report this heterogeneity explicitly and offers methodological recommendations for studies relying on long-delayed recall.

Keywords: case study, higher education, recall validity, retrospective evaluation, sustainability competencies, transformative learning

INTRODUCTION

Sustainability education research has increasingly turned to graduates and alumni as sources of evidence for the long-term effects of courses and programs, on the reasoning that durable learning is precisely what students carry with them after a course has ended. Cheang et al. (2021) relied on this reasoning when evaluating a co-curricular ecogarden program well after it ended. Zhang and Gibson (2021) did the same in their study of former study-abroad participants, and Montes de Oca et al. (2026) followed a similar logic in their account of a university sustainability certificate program. This turn toward retrospective accounts is understandable. Few institutions track competency development longitudinally, and alumni reflection is often the only available window on whether sustainability learning persists beyond the final exam. Yet retrospective self-report carries well-documented risks. Autobiographical memory is reconstructive rather than archival, and reports about extended past periods are shaped by the structure of memory itself rather than by stable, retrievable records of what happened (Schwarz, 2007). Bell and Bell (2018), tracking the same families over a twenty-five-year interval, found that retrospective and prospective reports of the same environment correlated only moderately (r = .30 to .45), and that retrospective accuracy varied by the type of content being recalled. These findings caution against treating long-delayed recall as a transparent record of past experience, a caution that applies directly to sustainability education research built on alumni reflection.

This article addresses that caution directly, using a case in which the gap between original course delivery and data collection was explicit, spanning four years. The course in question, "Sustainability and Development in Latin America: Past, Present, Future" (7.5 credits, Uppsala University, course code UU-19514), ran in autumn 2021 at the Centre for Environment and Development Studies (CEMUS) and was discontinued a few years later owing to funding constraints and low enrollment, not pedagogical failure. It combined Freirean problem-posing pedagogy with structured small-group discussion across three blocks per session, and it was the empirical case examined in a master's thesis on pedagogical innovation in sustainability education (Dutra, 2025). That thesis collected ten semi-structured interviews in 2025 with

This article is derived from data and analysis originally collected for the author's master's thesis, "Pedagogical innovations in sustainability education – A case study" (course code EX0901, Swedish University of Agricultural Sciences, 2025).

Copyright © 2026 by Author/s and Licensed by Modestum DOO, Serbia. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

five students, three instructors and two staff members, all reflecting on a course that had ended roughly four years before they were asked about it.

The original thesis used these interviews, together with course documents and 2021 evaluation data, to assess how well the course integrated five sustainability competencies identified by Wiek et al. (2011), namely systems thinking, anticipatory, normative, strategic and interpersonal competence. That analysis produced a synthesized account in which interpersonal and normative competence were reported as well integrated and anticipatory and strategic competence as comparatively weak. The present article does not repeat that synthesis. Instead, it returns to the same interview material with a different question. Rather than asking what competencies the course built, it asks what the act of recalling those competencies four years later actually looked like, and what that recall process itself can tell sustainability education research about the evidentiary status of retrospective alumni accounts.

Two patterns in the data motivate this reframing. First, participants repeatedly and independently reported difficulty distinguishing this course from other, similarly structured courses they had taken in the same period, a finding the original thesis noted only as a brief methodological limitation. Second, individual ratings of specific competencies, particularly anticipatory and strategic competence, varied considerably across participants in ways that did not collapse neatly into the thesis's "strong on three, weak on two" pattern. Some participants rated these competencies as well integrated, others rated them poorly, often in the same interview in which they expressed uncertainty about which course they were describing. Read together, these two patterns suggest that the curated, synthesized narrative common to single-case retrospective studies can smooth over a more heterogeneous and, this article argues, more informative underlying reality.

The article makes two contributions. Substantively, it offers a case-based account of how sustainability competencies are recollected, rather than measured, several years after a course ends, set against a contemporaneous 2021 evaluation baseline that allows some comparison between in-course and retrospective accounts. Methodologically, it argues that single-case retrospective designs in sustainability education research should treat recall heterogeneity and conflation as reportable findings rather than as noise to be resolved through synthesis, and it offers concrete recommendations for how future studies might document this heterogeneity rather than narrate around it. The study is explicitly not a longitudinal design in the conventional sense. There is no 2021 measurement of individual competencies against which the 2025 self-assessments can be formally compared, only a 2021 course evaluation describing overall satisfaction and thematic strengths and weaknesses. This limitation is treated throughout as part of the object of study rather than as a caveat to be minimized.

LITERATURE REVIEW

This review situates the article's empirical puzzle within four related literatures. It begins with what is known about the structure and reliability of retrospective self-report in general, before turning to the sustainability competency framework the original thesis used to evaluate the course, the pedagogical theories that predicted durable learning from this particular course design, and prior alumni studies in sustainability education that have relied on retrospective accounts without examining their reliability. The section closes by drawing these strands together into the specific gap the remainder of the article addresses, namely that recall heterogeneity has rarely been treated as a finding in its own right.

Retrospective Self-Report and the Structure of Memory

Retrospective self-report has a long history of methodological scrutiny outside education research, and that scrutiny is directly relevant to how sustainability education studies use alumni accounts. Schwarz (2007) distinguishes between historical information, which respondents can only access through memory because no alternative record exists, and behavior or experience that could in principle be captured in real time. He argues that autobiographical memory is organized hierarchically, with extended periods nested above specific events, and that this structure systematically biases recall toward general impressions and away from discrete episodes as the interval since the event lengthens. Henry et al. (1994) supply the most direct empirical test of this concern relevant to the present case. Comparing prospective data collected since birth from a New Zealand birth cohort with retrospective reports the same individuals gave at age eighteen, across seven content domains, they found that psychosocial variables, including subjective psychological states and family processes, showed the lowest agreement between prospective and retrospective measures of any domain tested, and that even statistically significant correlations corresponded to poor absolute agreement. Their conclusion, that retrospective reports should be used only to establish the relative standing of individuals within a distribution rather than to estimate precise frequencies or dates, applies with some force to any study asking participants to rate, on a numerical scale, how well a specific competency was integrated into a course several years earlier. Bell and Bell (2018) reached a similar conclusion over an even longer interval, twenty-five years, finding correlations of .30 to .45 between prospective and retrospective reports of family environment, with retrospective accuracy higher for emotionally salient content than for more cognitively detailed content.

A related but distinct strand of evidence concerns the decay of learned material itself, independent of any distortion in how that material is later reported. Murre and Dros (2015), replicating Ebbinghaus' classic 19th century forgetting-curve experiment, confirmed a steep early decline in retained material followed by a long, slow tail, with the overall shape of the curve robust across replications despite individual variability. This literature speaks to a different mechanism than Schwarz (2007) or Henry et al. (1994). Where the latter concern the fidelity of self-report about a past state, the

forgetting-curve literature concerns the loss of the underlying content itself. Both mechanisms are plausible candidates for what the present case observes, and the two are not easily disentangled using interview data alone, a point returned to in the discussion.

Sustainability Competencies as an Evaluative Framework

The competency framework most widely used to evaluate sustainability course outcomes identifies systems thinking, anticipatory, normative, strategic and interpersonal competence as the core capacities such education should build (Wiek et al., 2011). The framework was developed as a literature synthesis intended to guide program design and evaluation, not as a recall-validated measurement instrument, and Wiek et al. (2011) are explicit that translating the framework into assessment remains an open challenge. Venn et al. (2024) extend the framework empirically by examining how experienced sustainability professionals, with an average of twenty-four years of work experience, describe the competencies they actually use, finding a cluster of intervention-oriented competencies bound together by an integrated learning competency that fuses topical knowledge with lived experience. Their concurrent, practice-based design indicates that competency self-description, even among highly experienced professionals, is articulated through accumulated practice rather than recalled as discrete course content, a pattern with direct implications for how any single-course alumni study should expect competency-related recall to behave several years after the fact.

Pedagogical Grounds for Expecting Durable Learning

The course examined in this article was designed around Freirean problem-posing pedagogy, in which education proceeds through dialogue between teacher and student rather than the one-directional transfer of fixed content, with the explicit aim of building critical consciousness that persists beyond the classroom (Freire, 2000). His later writing develops the same commitment to dialogic, consciousness-raising education (Freire, 2005), and his account of liberatory pedagogy as a practice of culture and power reinforces the same point (Freire, 1985). Mezirow's (1997) theory of transformative learning, which describes learning as a process of revising the frames of reference through which adults interpret experience, offers a complementary rationale. A course designed to provoke this kind of reframing should, in principle, produce change that participants can recognize and describe years later, precisely because the change concerns their frame of reference rather than a discrete fact they might forget. These two literatures matter to the present article not as competing explanations to be tested, but because they raise the stakes of the recall problem identified here. If a course is designed, on explicit pedagogical grounds, to produce durable, frame-level change, then evidence that participants nonetheless struggle to distinguish this course from similar courses or disagree sharply with one another about which competencies it built, is harder to dismiss as a trivial artifact of ordinary forgetting. It suggests instead that even pedagogically ambitious, theoretically well-grounded courses may leave behind a retrospective record that is considerably messier than transformative learning theory alone would predict.

Retrospective and Alumni Studies in Sustainability Education

A small number of studies have examined sustainability or environmental learning outcomes years after a program ended. Zhang and Gibson (2021) interviewed thirty-one former study-abroad participants, aged twenty to thirty-three, about the long-term effects of a short-term sustainability-themed program, identifying lifestyle, professional development and responsible travel behavior as durable outcomes attributed to the program. Cheang et al. (2021) used a phenomenographic, retrospective evaluation of a co-curricular ecogarden program in Hong Kong, drawing on a serial time-point sampling strategy across four cohorts to assess the validity of retrospective evaluation itself, finding that the depth of participants' reflection increased with time since participation. Montes de Oca et al. (2026) similarly relied on alumni reflection from a university sustainability certificate program to identify experiential learning, interdisciplinary pedagogy, institutional support and career preparation as features alumni associated with lasting impact, while explicitly flagging their retrospective, self-selected design as a limit on causal claims. None of these three studies treats the accuracy or internal coherence of recall as part of its findings. Each uses retrospective report as a relatively transparent proxy for durable learning, a working assumption the memory literature reviewed above gives some reason to question.

Institutional Context and the Limits of Pedagogical Persistence

The course's discontinuation after a single funding-constrained, low-enrollment iteration is relevant to this article for a reason distinct from institutional survival itself. Rogers' (1983) diffusion of innovations framework holds that a practice's persistence depends on sustained adoption rather than on demonstrated effectiveness, and Lozano et al. (2015), surveying sustainability implementation across higher education institutions worldwide, found that resourcing and institutional commitment, rather than pedagogical design, were the recurrent determinants of whether such initiatives endured. Read against the present case, this matters because it sharpens what is actually being asked of participants' memory. They are not being asked to recall an ongoing, institutionally reinforced practice that continued to shape their environment after the course ended. Instead, they are being asked to recall a single, time-bounded iteration of a course that no longer exists, was never repeated, and had no institutional afterlife to anchor or refresh their memory of it. Where Cheang et al.'s (2021) ecogarden program persisted across multiple cohorts for years, giving later participants institutional continuity to draw on, this course offered no such continuity. Its discontinuation is therefore treated here not as a finding about institutional fragility in its own right, a question this case has been used to examine elsewhere, but as a condition that plausibly intensifies the recall difficulties this article documents. A discontinued, unrepeated course may be a harder thing to remember accurately than a program that continues to exist as a point of reference.

Synthesis and the Gap Addressed

Two further sources frame the methodological landscape this article speaks to. Özata Yücel and Özkan (2018) tracked the same fifty-five prospective science teachers across a genuine longitudinal design, testing their environmental perceptions at two points three years apart, and found change that was real but uneven, illustrating what a true longitudinal design looks like and underscoring what is given up methodologically when only a single retrospective measurement is available, as in the present case. Fróna (2026), reviewing sixty-three review-level studies on sustainability knowledge transfer in higher education, concludes that the field lacks clear, consistently documented links between specific pedagogical mechanisms and measurable student outcomes, a gap attributable partly to fragmentation across disciplinary literatures.

Taken together, this literature establishes the gap this article addresses. Sustainability education research increasingly depends on retrospective alumni accounts to claim durable impact. Zhang and Gibson (2021) drew this conclusion from former study-abroad participants, Cheang et al. (2021) from an ecogarden program, and Montes de Oca et al. (2026) from a sustainability certificate program. This dependence on alumni accounts is set against a pedagogical backdrop that gives strong theoretical reasons to expect such impact to be durable and self-recognizable, articulated in Freire's (2000) account of consciousness-raising education and in Mezirow's (1997) theory of transformative learning. Yet this research rarely engages with the substantial evidence from memory research that retrospective accounts over long intervals reorganize, generalize and sometimes conflate distinct experiences. Schwarz (2007) demonstrated this reorganization in his hierarchical model of autobiographical memory, Henry et al. (1994) confirmed it empirically in a longitudinal birth-cohort study, and Bell and Bell (2018) found the same pattern over a twenty-five-year interval. Nor does this literature distinguish that distortion of report from the separate, well-documented decay of the underlying material itself, a process Murre and Dros (2015) traced directly. Where retrospective sustainability education studies do examine their own method, they tend to treat increasing reflective depth as evidence of validity, as Cheang et al. (2021) did, without testing whether that depth coexists with factual conflation between similar experiences. This is a question the present case's interview data raises directly, and the following Materials and Methods section is designed to examine it.

MATERIALS AND METHODS

This section sets out how that gap is investigated empirically. It describes the single-case design and its explicit departure from a longitudinal logic, the two bodies of data drawn on, the inductive analytic approach applied to the interview material, the steps taken to support the trustworthiness of the analysis, and the ethical basis for reusing data originally collected for the author's thesis.

Design

This article adopts a qualitative, single-case study design (Yin, 2018), using the same bounded case examined in the originating thesis, namely the 2021 iteration of "Sustainability and Development in Latin America: Past, Present, Future" at Uppsala University. A single-case design is appropriate here because the object of analysis is not generalizable prevalence but the detailed, contextualized account of how one bounded group of participants recalled one bounded educational experience (Yin, 2018). Unlike the original thesis, which used this case to evaluate pedagogical innovation, the present article reanalyzes the same primary data with a different research question, treating the interviews not as a transparent record of what the course achieved but as an object of study in their own right. They are treated here as an instance of retrospective self-report collected four years after the experience it describes. The design is explicitly not longitudinal in the conventional sense. A true longitudinal design would require measuring the same competencies at two points in time, once during or immediately after the course and once in 2025 (cf. Özata Yücel & Özkan, 2018, who tested the same individuals twice across three years). No such repeated individual measurement exists here. What exists instead is a 2021 course evaluation describing aggregate satisfaction and thematically coded strengths and weaknesses, together with a 2025 set of individual retrospective interviews. The comparison between them is treated throughout as a contrast between two different kinds of evidence collected four years apart, not as a before-and-after measurement of the same variable.

Data Sources

Two bodies of data anchor the analysis. The first is the 2021 contemporaneous evaluation record, comprising the course's short report and long report, prepared by the course coordinator and the CEMUS educational coordinator, and the full course evaluation survey administered by Uppsala University, completed by 18 of 19 active students (94.74% response rate), reporting an overall course satisfaction score of 4.7 out of 5. This record provides a baseline characterization of the course as experienced at the time, generated independently of the 2025 interviews and not subject to four years of recall. The second body of data is the set of ten semi-structured interviews conducted in March and April 2025, comprising five former students (student 001 to student 005), three instructors (instructor 001, instructor 002, and instructor 003) and two staff members (staff 001 and a second staff member, the course's former director of studies, referred to here as staff 002). All ten interviews followed a common guide covering participant background, numerical ratings (1 to 5) of how well five sustainability competencies were integrated into the course, open questions on transformative learning and student-led pedagogy, questions on institutional and organizational support, a dedicated section on longitudinal perspective and long-term impact, and closing reflections. This article draws on the interview transcripts and the summaries prepared from them for participant validation, across all ten interviews.

Analytic Approach

The interview material was analyzed using thematic analysis (Braun & Clarke, 2006), proceeding inductively rather than testing a predefined coding frame. Two analytic passes were made across the material. The first pass coded explicit statements of recall difficulty, uncertainty or conflation between this course and other similar courses, wherever participants used language indicating they could not confidently isolate this course from comparable experiences. The second pass extracted each participant's numerical ratings of the five Wiek et al. (2011) competencies and compared them across participants, without first averaging or synthesizing them into a single pattern, in order to examine the degree of agreement or disagreement in the ratings themselves. Both passes draw on the same primary data used in the originating thesis, but the coding categories used here, recall difficulty and inter-rater heterogeneity, are specific to this article and were not the categories the thesis applied to this material.

Trustworthiness

Following Lincoln and Guba (1985), credibility is supported through triangulation between the 2021 evaluation record and the 2025 interviews. Dependability and confirmability are supported by grounding every claim in this article in material the reader can in principle trace back to a specific interview or document. No claim is made about any participant's views beyond what is recorded in these sources. Transferability is necessarily limited, consistent with Lincoln and Guba's (1985) argument that naturalistic case studies generalize only through the reader's own judgment about the similarity of the case to other contexts, and this limitation is treated as a finding of the design rather than a defect to be corrected. Coding was conducted by a single analyst. The absence of a second coder for formal inter-rater reliability checking is acknowledged as a limitation in the Limitations section.

Ethical Considerations

All participants were provided with detailed information about the study's aims, methods and their rights prior to giving informed consent, and all interview data were anonymized using the participant coding convention described above (student 001 to student 005, instructor 001 to instructor 003, staff 001 and staff 002). Because the original researcher served as both course facilitator and thesis investigator, the source study mitigated this conflict of interest through peer debriefing sessions and independent external review of the course design and implementation (Dutra, 2025). This article relies on the same anonymized, consented dataset and introduces no new data collection. No further ethical approval was therefore required beyond that obtained for the original thesis. With the design, data and analytic approach established, the next section turns to what the interviews actually show.

RESULTS

The Results are presented in three steps. The section first documents recall conflation as a pattern that recurs independently across participants, then sets out the heterogeneity in their numerical competency ratings and finally contrasts both with the more uniform picture given by the 2021 contemporaneous evaluation.

Recall Conflation as a Patterned, Cross-Participant Phenomenon

Across the ten interviews, references to difficulty isolating this course from other, similarly structured courses appear independently among participants with no apparent reason to coordinate their accounts. Student 004 was recalled by the interview summary as having "struggled to remember every precise example due to the time that had passed," despite rating most competencies highly. Student 005's account shows the same pattern more explicitly. When discussing anticipatory competence, the participant "reaffirmed its presence while clarifying that their memory intertwined this course with others taken concurrently" (student 005 interview summary, 2025). The clearest instance comes from the one interview from staff 002, who repeatedly qualified their own answers with phrases such as "I would speculate," "I don't know," and, when asked to confirm a detail about whether the 2021 course had run online, "could you repeat the question then because then my mind drifted off." This is not a single participant's idiosyncrasy. It recurs across a student, a staff member, and, in less explicit form, several instructors who hedged ratings with phrases such as "I don't fully know" or "that would also be a bit hard to say specifically for this course."

Heterogeneity in Competency Ratings

Where numerical ratings of the five Wiek et al. (2011) competencies were recorded across interviews, the pattern across participants is considerably less uniform than the thesis's synthesized account of "strong interpersonal and normative competence, weak anticipatory and strategic competence" suggests. The ratings, recovered across all ten interviews, are reported in full in **Table 1**.

Two patterns stand out. Systems thinking attracted close to unanimous agreement, clustering at 4 with isolated ratings of 3 and 5. Interpersonal collaboration showed a related but not identical pattern. Eight participants reported this competency only as a range of 4 to 5. Student 001 rated it a 3, the one score below 4 recorded for this item across all ten interviews. Anticipatory competence and strategic planning, by contrast, showed considerably more spread, ranging from 3 to 5 and from 2 to 4, respectively. This spread shows that the thesis's binary "weak" characterization of these two competencies reflects a synthesizing choice rather than a description of consensus in the underlying data. Student 003 and instructor 002 illustrate this directly. They rated strategic planning at opposite ends of the scale, 2 and 4, respectively, while describing structurally similar features of the course. Student 003 cited the absence of "practical case study" engagement, and instructor 002 cited a "very clear sort of sense of how these building blocks were building to the final assignment." Both views are internally coherent. They simply disagree, and that disagreement is itself a finding the original synthesis did not preserve.

Table 1. Individual competency ratings reported across all ten interviews

Competency	I 001	I 002	I 003	SF 001	SF 002	S 001	S 002	S 003	S 004	S 005
Systems thinking	4	5	4	4	3	-[4]	5	4	4	4
Anticipatory competence	-[1]	3	3	4	4	4	4	4	5	4
Normative or ethical judgment	5	2-3	5	5	4	4	5	5	5	3
Strategic planning	4	4	3	-[2]	3	2	4	2	4	3
Interpersonal collaboration	4-5[3]	4-5[3]	4-5[3]	4-5[3]	4-5[3]	3	5	4-5[3]	4-5[3]	4-5[3]

Note. I: Instructor; SF: Staff; S: Student; Ratings are on a 1-to-5 scale; All ratings are drawn from interview transcripts and summaries prepared for participant validation; [1]Instructor 001 did not assign a number for anticipatory competence, noting their participation "was mostly about historical and current perspectives, not future scenarios"; [2]Staff 001 did not provide a rating for strategic planning, stating they "could not comment in detail"; [3]For eight of the ten participants, interpersonal collaboration was reported in the underlying interview transcripts and summaries only as a range, from 4 to 5, without individual figures being broken out; Student 001 and student 002 gave individual ratings of 3 and 5, respectively; [4]Student 001 did not state a numeric rating for systems thinking; They instead described at length how the course reshaped their understanding of structural inequality in Latin America without assigning the competency a number

Contrast with the 2021 Contemporaneous Baseline

The 2021 evaluation record offers a different kind of evidence than the 2025 interviews, and the comparison must be read cautiously. The 2021 survey did not ask students to rate the five Wiek et al. (2011) competencies individually, so no direct, like-for-like comparison between 2021 and 2025 scores is possible. What the 2021 record does provide is an aggregate satisfaction figure (4.7 of 5, 94% overall impression, based on an 18-of-19 response rate) and a thematically consistent list of student-identified strengths (small-group discussion, student-led seminars, and engagement of the course coordinator) and weaknesses (literature diversity, structural clarity, and time constraints). This aggregate picture is considerably more uniform than the 2025 interview data. Where the 2021 survey shows broad, low-variance agreement on what worked and what did not, the 2025 interviews show considerable disagreement on the specific competencies the same course is said to have built. One plausible reading is that the 2021 instrument, being aggregated across 18 respondents and reported only as summary statistics and thematic counts, would have smoothed over individual disagreement in the same way the thesis's later synthesis did, and that the heterogeneity visible in the 2025 data was always present but never visible in a format that could surface it. This reading cannot be confirmed without access to individual-level 2021 responses, which were not collected in a form permitting this comparison, and is offered here as a limitation of the available baseline rather than a finding. Read together, these findings raise an interpretive question taken up in the Discussion, namely whether this heterogeneity reflects something more than ordinary variability in subjective ratings.

DISCUSSION

The discussion considers what these patterns mean. It first asks whether the heterogeneity reported in the Results could simply be ordinary measurement noise, then proposes a memory-based explanation for why this particular course may have been difficult to recall accurately, traces a pedagogical irony this explanation implies, draws out methodological recommendations for future retrospective studies, and closes by stating the study's limitations plainly.

Is This Heterogeneity, or Just Noise?

A reasonable objection to the results is that ratings from ten people on a five-point scale will show some spread under almost any conditions, retrospective or not, and that a 2-to-4 range on strategic planning is not obviously different from what a same-day, in-course survey of ten people might also produce. This objection has force and should not be dismissed. What distinguishes the present case from ordinary subjective variance is not the spread itself but its pairing with explicit, independent statements of recall uncertainty from several of the same participants who gave divergent ratings. Student 003 rated strategic planning a 2 while explicitly wanting more "practical case study" engagement, a substantive, specific objection rather than a vague hedge. Instructor 002 rated the same competency a 4 while describing concrete structural features ("how these building blocks were building to the final assignment"). Neither account reads as a guess. The disagreement is between two specific, confidently held, but incompatible characterizations of the same course, which is a different phenomenon from random measurement noise, though this article cannot rule out noise as a contributing factor, only argue that the qualitative content of the interviews makes pure noise an incomplete explanation.

Why This Course May Be Hard to Remember Accurately

Schwarz (2007) argues that autobiographical memory is organized hierarchically, with broad categories of experience nested above specific events. This structure predicts exactly the kind of conflation observed here, but only under a particular condition, namely that the specific event in question closely resembles other events filed under the same broad category. That condition holds strongly in this case. Multiple participants independently noted that CEMUS courses in general share a similar pedagogical architecture, including flipped-classroom delivery, circle-based discussion and student-led facilitation, which means that "this course" was, for several participants, one instance among several near-identical instances filed under a single higher-order memory category ("CEMUS courses I took around that time"), rather than a distinctive, easily isolated episode. Staff 002 confirmed directly that this pedagogical format was not unique to the course under study but characteristic of CEMUS courses generally. Henry et al. (1994) found that psychosocial variables, precisely the kind of content sustainability competency self-ratings require, showed the weakest prospective-retrospective agreement of any domain they tested. The present case suggests a mechanism for why. Psychosocial impressions of a course are more easily merged with impressions of similar courses than factual, dateable details would be.

This also clarifies a distinction that the Results left open. The recall difficulty documented here looks less like the simple

decay described by the Ebbinghaus forgetting curve (Murre & Dros, 2015), in which specific content is progressively lost, and more like the reconstructive misattribution described by Schwarz (2007), in which content is retained but assigned to the wrong source. Participants do not generally report a blank where the course should be. They report vivid, specific memories, such as movie nights, circle seating and particular guest lecturers, that they then struggle to attribute confidently to this course rather than to a structurally similar one. This is a meaningfully different problem for retrospective sustainability education research than ordinary forgetting, and it has not, to this article's knowledge, been previously documented in this literature.

A Pedagogical Irony

This produces a genuine tension for the pedagogical literature the course's design drew on. Freirean problem-posing pedagogy and Mezirow's (1997) transformative learning theory both predict that dialogic, reflective, student-centered teaching should produce durable, frame-level change precisely because it engages participants more deeply than transmission-style teaching. Nothing in the present data contradicts that prediction at the level of content. Participants describe lasting shifts in how they think about development, justice and their own academic orientation (student 001 and student 002), and these accounts are often the most fluent and confident parts of the interviews. The irony is that the very pedagogical consistency that CEMUS cultivates across its courses, in order to reliably produce this kind of transformative engagement, is plausibly also what erodes participants' ability to attribute that change to any one course four years later. A teaching model designed for depth of impact may, by being applied consistently across many similar courses, undermine the episodic distinctiveness that would let alumni later trace that impact back to a specific course. This suggests that transformative learning and accurate retrospective attribution may be in some tension with one another, a possibility the transformative learning literature, focused as it is on whether change occurs rather than on whether participants can later locate its source, does not currently address.

How often students confuse one course with another cannot be estimated from ten interviews, and the education literature offers no direct figure. Memory research nonetheless treats such confusion as expected rather than anomalous. The more similar two sources are, the harder it becomes to attribute a memory to the correct one, and people may recognize material almost perfectly while identifying its source no better than chance (Johnson et al., 1993). Retention studies of formal coursework also suggest that only around 30% of course knowledge remains after four years (Custers, 2010). What survives is less the course as an episode than the understanding it left behind. Conway et al. (1997) documented this shift from remembering to knowing among undergraduates and regarded it as precisely the change educators should hope to see in their students.

The interviews show this in varying degrees. Student 005 and staff 002 could not reliably separate this course from other CEMUS courses, whereas student 001 kept three student-led courses clearly apart, and instructor 001, who had taught few sustainability courses since, recalled the two guest sessions "as something very different." Student 002 had "images in my head and also the topics, but I don't connect both of them" yet retained "a grid to analyze things and to link sustainability with development," which they continued to apply to news and political questions. Barber (2012) calls this integration of learning, whose most developed form, synthesis, combines two or more insights into new understanding, and found that students largely achieved it on their own, without a mentor or guide. Conflation is therefore a problem for evaluation but may, from an educational standpoint, signal success. It is not, however, evidence of integration in itself, and the present data show integration clearly in one case only.

Methodological Implications

Three recommendations follow for single-case retrospective designs in sustainability education research. First, individual-level ratings or assessments should be reported alongside any synthesized narrative, not folded into it, so that heterogeneity of the kind documented here remains visible to readers rather than disappearing into a single pattern. Second, interview protocols for retrospective alumni studies should explicitly ask participants whether they can distinguish the target course from other similar experiences, treating conflation as a data point to be recorded rather than a contaminant to be edited around. Third, where an institution offers a pedagogically consistent suite of courses, as CEMUS does, retrospective studies of any single course within that suite should treat the suite's consistency itself as a methodological variable likely to affect recall, rather than assuming each course will be remembered as a discrete unit.

Limitations

This study's limitations are substantial and are stated plainly rather than minimized. The sample is small, ten interviews, all ten of which ultimately yielded numerical competency ratings once raw transcripts were checked against prepared summaries, though four participants left at least one competency unrated or rated only in a non-numeric form, and self-selected, drawn from those willing to be re-contacted four years after a discontinued course, which may bias the sample toward participants with stronger or more coherent memories than non-respondents would have. Coding was conducted by a single analyst without a second coder for inter-rater reliability, consistent with the original thesis's approach but a limitation Lincoln and Guba (1985) would flag for confirmability. The 2021 baseline, as noted in the Results, does not permit individual-level comparison, only an aggregate one, which weakens the longitudinal contrast this article can draw. Finally, this case is a single, idiosyncratic course at a single institution, and the pedagogical-consistency mechanism proposed above is offered as a plausible, case-grounded explanation, not a tested or generalizable claim.

CONCLUSIONS

This article asked what happens when sustainability competency self-assessment is collected four years after a course ends, rather than treating that delay as a regrettable but unavoidable feature of alumni research. The answer, drawn

from ten interviews revisiting a discontinued course at Uppsala University, is that the delay does not simply attenuate recall. It reorganizes it. Participants did not report a faded version of the original thesis's synthesized narrative. They reported specific, often vivid, but inconsistently attributed memories, in several cases unable to confidently separate this course from structurally similar courses taken in the same period. This article has argued that this conflation is itself diagnostic. It is consistent with Schwarz's (2007) hierarchical model of autobiographical memory, under which highly similar experiences sharing a pedagogical format are filed under a common category and become difficult to identify in recall, and it offers one plausible explanation for why psychosocial content shows the weak prospective-retrospective agreement that population-scale studies have documented. Henry et al. (1994) reported this weak agreement directly, and Bell and Bell (2018) found the same pattern over a longer interval. The pedagogical consistency that CEMUS cultivates across its course offerings, designed in the Freirean and transformative-learning traditions to produce durable change, may be precisely what makes any single course within that offering hard to recall as a distinct episode four years on. This is offered as a case-grounded hypothesis, not a generalizable finding, but it is, to this article's knowledge, a connection not previously drawn in the sustainability education literature.

The practical implication is not that retrospective alumni research should be abandoned, but that it should report what it actually finds rather than what a clean narrative would prefer. Single-case studies relying on long-delayed recall should disaggregate individual responses, treat conflation and recall uncertainty as data rather than noise, and, where the institution offers a suite of pedagogically similar courses, treat that similarity itself as a factor likely to shape what alumni can and cannot reliably attribute to the course under study.

Funding: No funding source is reported for this study.

Ethical statement: This study reuses anonymized, previously consented interview data from the original thesis study. No new human-subjects data were collected, and no additional ethical approval was required.

AI statement: Generative AI (Claude by Anthropic) was used to assist only with language/grammar editing under the author's direction. The author verified all empirical claims, citations and interpretive arguments against primary sources and takes full responsibility for the final content.

Declaration of interest: No conflict of interest is declared by the author.

Data sharing statement: The anonymized interview data are not publicly available due to confidentiality commitments made to participants at the time of data collection.

REFERENCES

Barber, J. P. (2012). Integration of learning: A grounded theory analysis of college students' learning. *American Educational Research Journal, 49*(3), 590-617. https://doi.org/10.3102/0002831212437854

Bell, D. C., & Bell, L. G. (2018). Accuracy of retrospective reports of family environment. *Journal of Child and Family Studies, 27*(4), 1029-1040. https://doi.org/10.1007/s10826-017-0948-5

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Cheang, C.-C., Ng, W.-K., Wong, Y.-S. D., Li, W.-C., & Tsoi, K.-H. (2021). Planting a seed of experience: Long term effects of a co-curricular ecogarden-based program in higher education in Hong Kong. *Frontiers in Psychology, 11*, Article 583319. https://doi.org/10.3389/fpsyg.2020.583319

Conway, M. A., Gardiner, J. M., Perfect, T. J., Anderson, S. J., & Cohen, G. M. (1997). Changes in memory awareness during learning: The acquisition of knowledge by psychology undergraduates. *Journal of Experimental Psychology: General, 126*(4), 393-413. https://doi.org/10.1037/0096-3445.126.4.393

Custers, E. J. F. M. (2010). Long-term retention of basic science knowledge: A review study. *Advances in Health Sciences Education, 15*(1), 109-128. https://doi.org/10.1007/s10459-008-9101-y

Dutra, A. (2025). *Pedagogical innovations in sustainability education – A case study* [Master's thesis, Swedish University of Agricultural Sciences]. https://stud.epsilon.slu.se/21177/1/dutra-a-20250622.pdf

Freire, P. (1985). *The politics of education: Culture, power, and liberation* (D. Macedo, Trans.). Bergin & Garvey.

Freire, P. (2000). *Pedagogy of the oppressed*. Continuum.

Freire, P. (2005). *Education for critical consciousness*. Continuum.

Fróna, D. (2026). Sustainability knowledge transfer in higher education: A narrative review. *Education Sciences, 16*(4), Article 570. https://doi.org/10.3390/educsci16040570

Henry, B., Moffitt, T. E., Caspi, A., Langley, J., & Silva, P. A. (1994). On the "remembrance of things past": A longitudinal evaluation of the retrospective method. *Psychological Assessment, 6*(2), 92-101. https://doi.org/10.1037/1040-3590.6.2.92

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*(1), 3-28. https://doi.org/10.1037/0033-2909.114.1.3

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE.

Lozano, R., Ceulemans, K., Alonso-Almeida, M., Huisingh, D., Lozano, F. J., Waas, T., Lambrechts, W., Lukman, R., & Hugé, J. (2015). A review of commitment and implementation of sustainable development in higher education: Results from a worldwide survey. *Journal of Cleaner Production, 108*(Part A), 1-18. https://doi.org/10.1016/j.jclepro.2014.09.048

Mezirow, J. (1997). Transformative learning: Theory to practice. *New Directions for Adult and Continuing Education, 1997*(74), 5-12. https://doi.org/10.1002/ace.7401

Montes de Oca, L., Navarro, M., & Byers, T. (2026). Understanding what makes sustainability education effective: Insights from a university certificate program. *Cleaner Production Letters, 10*, Article 100133. https://doi.org/10.1016/j.clpl.2026.100133

Murre, J. M. J., & Dros, J. (2015). Replication and analysis of Ebbinghaus' forgetting curve. *PLoS ONE, 10*(7), Article e0120644. https://doi.org/10.1371/journal.pone.0120644

Özata Yücel, E., & Özkan, M. (2018). Analysis of change in the environmental perceptions of prospective science teachers: A longitudinal study. *Asia-Pacific Forum on Science Learning and Teaching, 19*(1), Article 3.

Rogers, E. M. (1983). *Diffusion of innovations* (3rd ed.). Free Press.

Schwarz, N. (2007). Retrospective and concurrent self-reports: The rationale for real-time data capture. In A. Stone, S. Shiffman, A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture: Self-reports in health research* (pp. 11-26). Oxford University Press. https://doi.org/10.1093/oso/9780195178715.003.0002

Venn, R., Vandenbussche, V., & Perez, P. (2024). Competencies of experienced sustainability professionals. *Frontiers in Sustainability, 5*, Article 1411041. https://doi.org/10.3389/frsus.2024.1411041

Wiek, A., Withycombe, L., & Redman, C. L. (2011). Key competencies in sustainability: A reference framework for academic program development. *Sustainability Science, 6*(2), 203-218. https://doi.org/10.1007/s11625-011-0132-6

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE.

Zhang, H., & Gibson, H. J. (2021). Long-term impact of study abroad on sustainability-related attitudes and behaviors. *Sustainability, 13*(4), Article 1953. https://doi.org/10.3390/su13041953