

PREVIEW-A&F: a mixed-methods eye-tracking methodology for prospective optimisation of electronic audit and feedback interventions

Marlena Klaic

`marlena.klaic@unimelb.edu.au`

The University of Melbourne

Amanda Douglass

The University of Melbourne

Damien Mannion

The University of Melbourne

Camille Paynter

The University of Melbourne

Viet Bui

The University of Melbourne

Jennifer Jones

The University of Melbourne

Lisa Guccione

Peter MacCallum Cancer Centre

Elise Mitri

Austin Health

Jason Trubiano

Austin Health

Andrew Anderson

The University of Melbourne

Method Article

Keywords: Electronic audit and feedback, eye-tracking, implementation science, behaviour change techniques, visual attention, digital health, mixed methods, intervention optimisation, Theoretical Domains Framework

Posted Date: August 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10353444/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Title: PREVIEW-A&F: a mixed-methods eye-tracking methodology for prospective optimisation of electronic audit and feedback interventions.

Short title: A methodology for eA&F optimisation

Authors

Marlena Klaic¹

Amanda Douglass²

Damien J Mannion³

Camille Paynter¹

Viet Bui¹

Jennifer Jones^{1,8}

Lisa Guccione^{6,7}

Elise Mitri^{4,5}

Jason A Trubiano^{4,5}

Andrew J Anderson²

Affiliations:

¹ School of Health Sciences, University of Melbourne, Melbourne, Australia

² Department of Optometry and Vision Sciences, University of Melbourne, Melbourne, Australia

³ Melbourne Data Analytics Platform, University of Melbourne, Melbourne, Australia

⁴ Department of Infectious Diseases and Immunology, Austin Health, Heidelberg, Australia

⁵ Department of Infectious Diseases, at the Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, Australia

⁶Department of Health Services Research, Peter MacCallum Cancer Centre, Melbourne, Australia

⁷Sir Peter MacCallum Department of Oncology, Faculty of Medicine, Dentistry and Health Science, The University of Melbourne, Melbourne, Australia

⁸Physiotherapy Department, Division of Allied Health, Austin Health, Heidelberg, Australia

Corresponding author: Marlena Klaic, PhD, School of Health Sciences, Level 7, Alan Gilbert Building, 161 Barry Street, The University of Melbourne, Victoria 3010, Australia.
marlena.klaic@unimelb.edu.au

Ethics approval: This study is a sub-study of the iNAAN project (HREC/71789/Austin-2021, approved by Austin Health HREC). Sub-study protocol number: 78719.

Grant Support: No external funding was received for this pilot study. Internal seed funding was provided by the Melbourne School of Health Sciences, University of Melbourne. Funding for the larger programme of research has been sought through an NHMRC Ideas Grant (Application ID: 2048067).

Commercial Relationships: None

Abstract

Background: Electronic audit and feedback is a widely used implementation strategy for improving clinician practice, yet interventions demonstrate variable effectiveness. Current approaches to electronic audit and feedback development rely heavily on subjective usability feedback and rarely include prospective optimisation prior to implementation. This may contribute to repeated redesign cycles when engagement barriers are only identified after deployment, representing an underexamined contributor to slower implementation pathways.

Methods: We developed and tested a theory-informed methodology for prospectively evaluating engagement with electronic audit and feedback interventions. Central to the methodology was an a priori framework linking each element in the tool to its intended behavioural change technique (BCT), hypothesised behavioural mechanism, and required level of visual attention. To test the methodology, health professionals enrolled as investigators in an international multi-site penicillin allergy implementation study (international Network of Antibiotic Allergy Nations) reviewed their site-specific electronic audit and feedback report while wearing an eye tracker and completing a subsequent think-aloud protocol. Gaze data and think-aloud data were independently analysed and then integrated to examine whether observed engagement aligned with the intervention's intended behavioural design.

Results: Integration of gaze and think-aloud data identified convergence and divergence between intended and observed engagement with BCT-linked interface elements. Of 29 areas of interest, 14 were Convergent (intended and observed engagement aligned across both data streams); one Hidden Opportunity (visually attended as intended but BCT activation incomplete); eight Attentional Mismatch (fixation duration deviated from the intended level despite adequate processing); three Divergent (engagement failure across both data streams); and three Silent Engagement. Joint interpretation of the two data streams generated targeted redesign recommendations that would not have been identifiable using either method alone.

Conclusions: This proof-of-concept study establishes the feasibility of a theory-informed methodology for prospectively evaluating engagement with electronic audit and feedback interventions prior to broader implementation. Integration of objective measures of overt visual attention with qualitative evidence of cognitive processing allows interface-level barriers to be

identified and addressed before implementation at scale, rather than after. Prospective optimisation may help reduce avoidable redesign cycles and support more efficient pathways to implementation readiness.

Contributions to the literature

- We describe a novel, theory-informed methodology that integrates behaviour change theory, eye-tracking, and think-aloud methods to prospectively optimise electronic audit and feedback (eA&F) interventions.
- The methodology uses a priori specification of behaviour change technique-linked interface elements and expected levels of overt visual attention to evaluate whether eA&F tools are visually engaged with as intended.
- By identifying engagement barriers prior to wider deployment, this methodology offers a structured approach for prospective optimisation that may reduce avoidable post-implementation redesign cycles.

Keywords

Electronic audit and feedback; eye-tracking; implementation science; behaviour change techniques; visual attention; digital health; mixed methods; intervention optimisation; Theoretical Domains Framework

Background

Audit and feedback (A&F) is an implementation strategy widely used in healthcare to improve clinician practice (1). The approach involves evaluating the performance of individuals or teams against a target, such as clinical guidelines, and providing those results back to clinicians. The mechanisms through which A&F changes behaviour are theoretically grounded: feedback activates knowledge of current performance, creates motivation to close a gap between actual and desired practice, and invokes subjective norms through social comparison (1-4). These mechanisms are typically operationalised through a range of behaviour change techniques (BCTs), including feedback on behaviour, discrepancy from goal and social comparison.

A large systematic review of 292 trials found that A&F leads to small-to-moderate improvements in clinician practice, with effects moderated by delivery features including individual-level data, peer benchmarking, action planning, and facilitation (1). More recently, the digitisation of healthcare systems has prompted a shift from paper-based A&F to electronic audit and feedback (eA&F). Clinical performance data are increasingly being delivered through digital interfaces such as automated dashboards and web-based reports (3, 5). eA&F tools can be embedded within routine digital infrastructures, enabling clinicians to access their performance data more frequently and potentially in near real time (5). However, realising this potential depends critically on whether clinicians engage with the interface in a way that activates the intended BCTs.

Despite increasing use of eA&F, effectiveness remains mixed. Research has found that less than half of measured outcomes (e.g., prescribing, guideline adherence) showed improvement following eA&F interventions (6, 7). The format of delivery, electronic versus paper-based, does not appear to influence the magnitude of the feedback effect (1, 6). Arguably, clinician interaction with the interface may be an important determinant of engagement. A targeted review of the eA&F literature conducted in April 2026 suggests that this variability in engagement is at least partly due to persistent gaps in how these tools are designed, specified and refined prior to deployment. Drawing on two existing systematic reviews (1, 8) and a sample of primary studies evaluating the development and outcomes of eA&F interventions, three key gaps were identified. First, explicit a priori specification of BCTs or mechanisms of action was uncommon. Studies where behavioural principles were described, most frequently social comparison, rarely reported

them as deliberate design decisions (9, 10). Second, few studies described iterative refinement of eA&F tools prior to large-scale deployment, with most progressing directly from development to effectiveness testing without a reported optimisation phase (11, 12). Third, where user engagement was evaluated, it typically consisted of self-report methods such as interviews, or post-hoc usability surveys (13, 14). No studies reported objective measures of visual attention or interaction behaviour. These findings suggest that eA&F tools are commonly treated as static interventions rather than iteratively refined behavioural systems. As a result, the field currently lacks methodological approaches for prospectively identifying and addressing engagement barriers during the design phase, which may contribute to repeated redesign cycles following poor uptake or weak engagement.

The subjective methods used to evaluate eA&F engagement have a further limitation that existing literature has not adequately addressed. Clinicians may not consciously recognise the barriers they encounter due to social desirability bias, habituation to stress, or limited awareness of their own information-processing difficulties (15-17). Consequently, subjective feedback alone may be insufficient to identify components of the tool that impact effective engagement. Understanding how clinicians visually engage with feedback interfaces is therefore a key methodological challenge for eA&F research. Attention determines which elements of a visual display receive cognitive processing and therefore influence decision making. One mechanism through which attention is directed is overt visual attention, the movement of the eyes to specific regions of a display, which can be measured objectively using eye-tracking technology. By defining areas of interest (AOIs) within a visual interface, eye-tracking software can quantify how frequently users fixate on specific elements and how long those fixations last. Within healthcare, eye-tracking has been applied across a range of contexts, including the study of cognitive overload in surgical performance and diagnostic decision making (18-20).

These challenges are also relevant to a broader question in implementation science: how can implementation strategies be optimised more efficiently? eA&F tools are commonly evaluated only after deployment, meaning engagement barriers may not become apparent until implementation is already underway. This typically results in either abandonment of the tool or repeated cycles of redesign, redeployment, and re-evaluation following poor uptake or weak engagement. The emerging concept of implementation pace focuses not only on whether

interventions are successfully implemented, but also on the efficiency with which they progress from design to sustainable use. One potential mechanism for improving implementation pace is the prospective identification and resolution of modifiable barriers before broader deployment. Rather than conceptualising implementation strategies as static products that are evaluated after release, prospective optimisation approaches treat them as iteratively refined behavioural systems whose design can be improved before large-scale implementation begins.

To address this gap, we developed and tested a theory-informed methodology, PREVIEW-A&F (Prospective Refinement using Eye-tracking and Verbal IntEgration Within electronic Audit & Feedback), that combines eye-tracking technology and think-aloud protocols to evaluate engagement with eA&F interventions. Central to the methodology is a theory-informed a priori framework linking each interface element to its intended BCT, hypothesised mechanism of action, and expected level of overt visual attention. By comparing observed gaze patterns with a priori expectations regarding intended engagement, and integrating these findings with think-aloud data, PREVIEW-A&F provides a structured approach for prospectively identifying interface-level barriers prior to broader implementation (Figure 1). Of note, the present methodology does not directly measure behaviour change, BCT activation, or implementation outcomes. Rather, it integrates objective evidence of overt visual attention with qualitative evidence of cognitive processing to determine whether observed engagement is consistent with the intended behavioural design of the intervention.

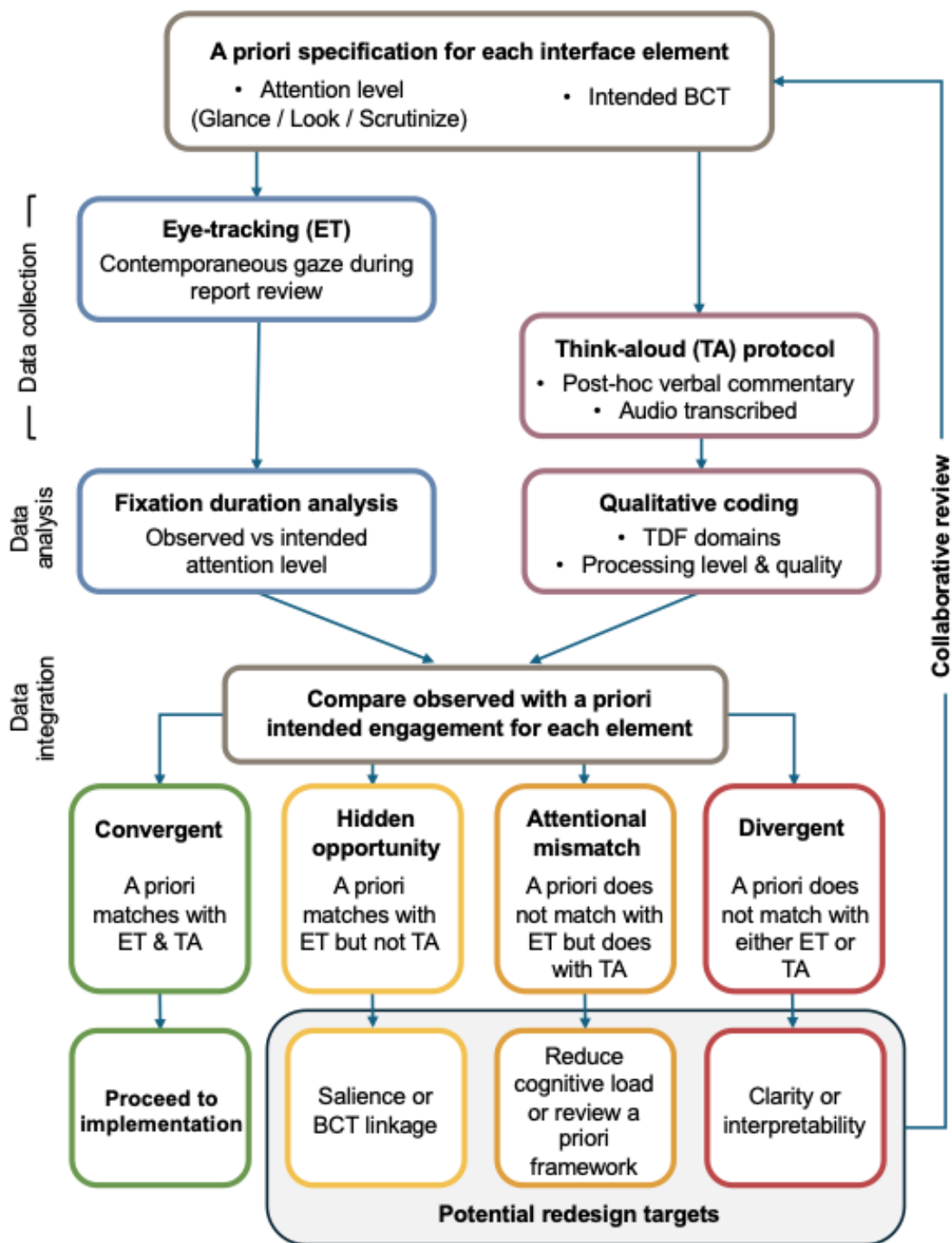


Figure 1 PREVIEW-A&F. A framework for prospective optimisation of digital health tools.

Overview of PREVIEW-A&F methodology

The PREVIEW-A&F methodology comprises five sequential phases applicable to any eA&F tool in any clinical domain, presented as a generalisable framework in Figure 1.

Phase 1 - Specify: before data collection, the research team establishes an a priori framework linking each interface element (AOI) to: (i) its intended BCT and hypothesised mechanism of action; (ii) its intended attention level i.e., Glance (rapid orientation; no BCT activation), Look (information extraction; contextual framing supporting adjacent BCT elements), or Scrutinise (sustained attention supporting BCT activation through comparison, reflection, and action planning); and (iii) the expected clinician action following full processing of the element. This specification serves as the benchmark against which observed engagement is later compared.

Phase 2 - Collect: participants review the eA&F tool while eye tracking records contemporaneous gaze behaviour. Immediately after, they complete a retrospective think-aloud protocol, providing a verbal commentary of their reasoning, interpretations, and any points of confusion. The two data streams are collected in the same session but remain independent until Phase 4.

Phase 3 - Analyse: each data stream is analysed independently. Fixation duration from the eye tracking data is examined using planned contrasts in a mixed model, comparing the total duration spent looking at each AOI (the fixation duration) against the group mean for its attention level; raw duration is used for Glance- and Look-classified AOIs, whereas fixation duration for Scrutinise-classified AOIs is normalised by the number of distinct information units to account for variation in information density. Think-aloud transcripts are coded using two complementary approaches: deductive coding against the Theoretical Domains Framework (TDF) (21) to identify which behavioural domains are activated, and inductive coding of processing level (Orienting, Extracting, Interpreting, Problem-solving) and processing quality (Successful, Mixed, Confused, Dismissed).

Phase 4 - Integrate: gaze and think-aloud data streams are brought together through collaborative multidisciplinary review. For each AOI, observed fixation duration is compared against the a priori intended attention level, and think-aloud coding is compared against the expected processing depth, classifying each AOI into one of four integration patterns (described in detail in the Integrated findings section below).

Phase 5 - Act: the review team examines each integration pattern and agrees on the appropriate redesign response. Convergent AOIs proceed to implementation. Each non-convergent pattern implies a distinct target: Hidden Opportunity → review BCT linkage or improve element salience; Attentional Mismatch → investigate why fixation duration deviates from the intended attention level; Divergent → redesign for clarity so the element is correctly understood. Agreed changes are made before re-evaluation. This iterative cycle continues until all BCT-critical AOIs achieve Convergent classification.

Study aims

The primary aim of this study was to develop and test the PREVIEW-A&F methodology for prospectively evaluating engagement with eA&F interventions using eye-tracking and think-aloud methods.

Secondary aims were to:

1. Examine whether observed patterns of overt visual attention aligned with a priori expectations regarding BCT-linked interface elements.
2. Demonstrate how integration of gaze and think-aloud data can identify interface-level barriers not detectable through subjective usability methods alone.
3. Illustrate how PREVIEW-A&F methodology may support prospective optimisation of eA&F interventions prior to broader implementation and so provide one potential mechanism for reducing downstream redesign requirements.

Methods

Study Design

The following sub-sections describe the application PREVIEW-A&F methodology to a case study: the iNAAN Health Service Report, an eA&F tool used in an international penicillin allergy management implementation study. This study used a convergent mixed-methods design, in which objective gaze data were collected alongside think-aloud verbal data. The two data streams were analysed independently and then integrated to identify convergences and

discrepancies. This study was conducted as a sub-study of the iNAAN project (HREC/71789) approved by Austin Health Human Research Ethics Committee, in accordance with the Declaration of Helsinki.

Participants and Setting

Participants were clinician investigators from participating sites of the iNAAN study, spanning a range of clinical backgrounds including infectious diseases clinicians, allergists, and pharmacists from Australian hospital sites. Recruitment took place at a professional symposium (the Drug Hypersensitivity Symposium, October 2024), where iNAAN investigators were invited to take part. Eligible participants were required to be currently enrolled as iNAAN investigators but did not need to have prior experience with the eA&F report. The study therefore included both participants familiar with the report and those viewing it for the first time (report-naïve participants). Informed e-consent was obtained prior to attendance at the symposium.

Electronic Audit and Feedback Report (Case Study)

The eA&F tool evaluated was the iNAAN Health Service Report, a two-page performance report provided to hospital clinical teams participating in the iNAAN study, designed to improve penicillin allergy documentation and de-labelling (i.e. removal of a patient's penicillin allergy following assessment and testing). The effectiveness and implementation outcomes of the broader iNAAN study have been reported elsewhere (22, 23). The iNAAN eA&F tool features key performance statistics with peer-comparison benchmarks, temporal trend charts, and documentation compliance indicators. Multiple iterations of the report had been developed and refined prior to this study through stakeholder input. As such, this work represents a concurrent application of PREVIEW-A&F methodology to an implementation strategy already in active use, with the aim of identifying opportunities for further optimisation. An example report is shown in Figure 2.

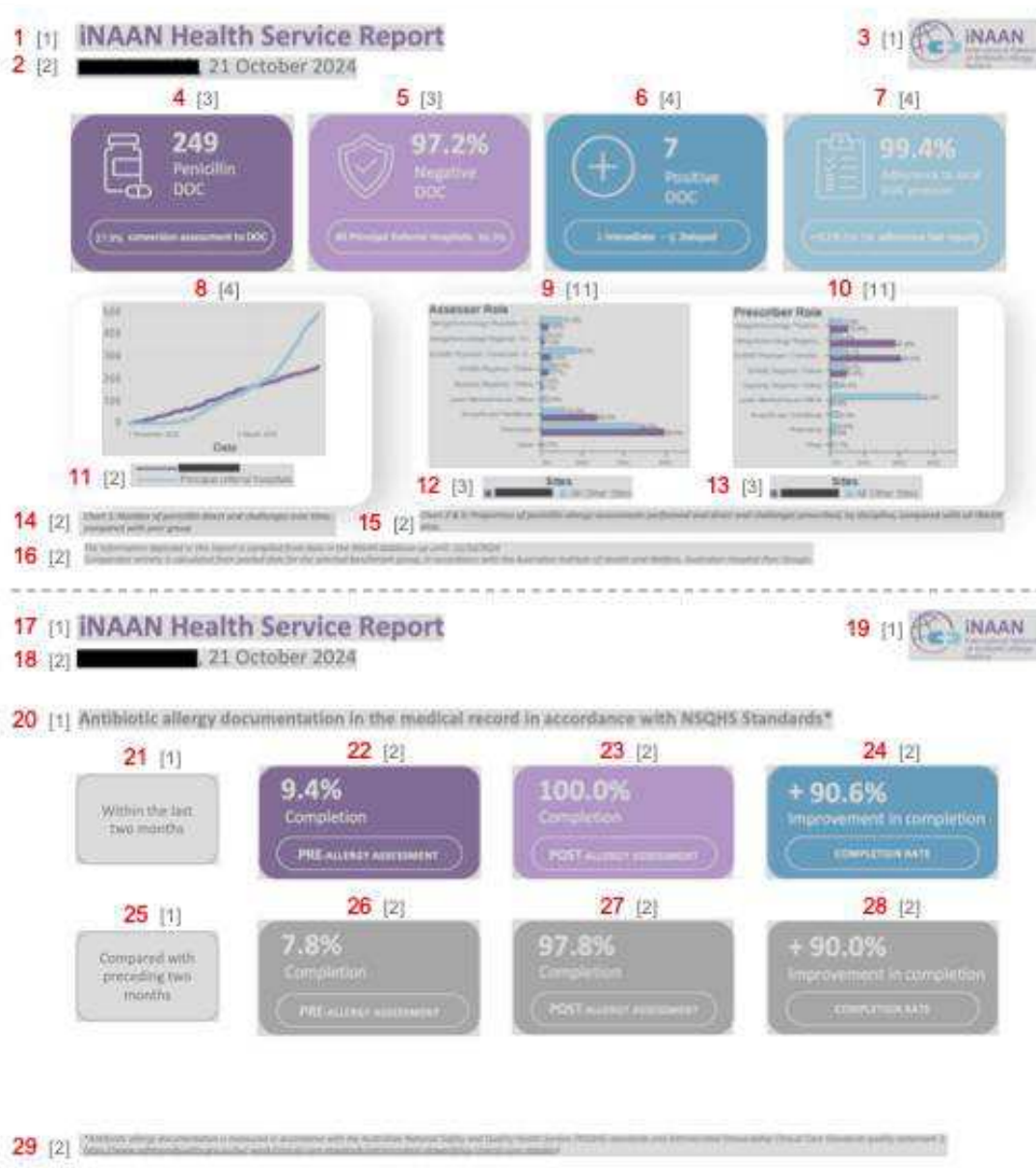


Figure 2 Example iNAAN report with Areas of Interest (AOIs) overlaid (grey rectangles). Numbering for the AOIs is given in red, with the subsequent grey number in parentheses indicating the a priori expectation of how many pieces of information needed to be extracted from the AOI. The report was delivered as a two-page pdf, with the page division indicated here by the horizontal dashed line. AOI rectangles and numbering were not visible in the report viewed by study participants. Black boxes have been used to redact identifying information and were also not present in the viewed report.

Area of Interest (AOI) Selection & Quantification: A Theoretically Grounded A Priori Framework

A central methodological contribution of this study is the a priori specification of Areas of Interest (AOIs). Throughout this paper, the terms 'interface element' and 'AOI' are used

interchangeably: an interface element refers to a discrete, bounded component of the eA&F tool (e.g., a chart, a metric tile, a legend). AOI is the eye-tracking operationalisation of that element, the precisely delimited region within the screen display to which fixations are mapped (Figure 2). In general, logos, titles, headers, footers and “infographic” elements providing a single piece of information were each assigned individual AOIs, whereas graphs were assigned AOIs for each of the figure panel, the figure key, and the figure legend.

We grounded the expected fixation duration for each AOI in the behaviour change theory underpinning A&F design. First, five of the researchers (AA, AD, MK, EM & JT) separately assessed how many pieces of information were contained in each AOI, with the final number then obtained through consensus (see Supplementary File 1).

Second, a multidisciplinary team comprising clinical content experts (JT and EM) and implementation scientists (MK, LG and VB) defined three elements for each AOI. Intended depth of attention was classified as Glance (rapid orientation; no BCT activation), Look (information extraction; contextual framing), or Scrutinise (sustained attention supporting comparative interpretation and BCT activation). Design purpose explicitly linked each element to its intended BCT(s). Expected subsequent action described what a clinician who had fully processed the item should do next. These categories are intended as heuristic attention classifications rather than direct measures of cognition or comprehension.

This framework serves two purposes. First, it enables hypothesis-driven analysis by comparing observed fixation duration with the pre-specified expectations regarding intended levels of overt visual attention, such that Scrutinise-classified elements are expected to attract proportionally greater fixation than orientation or contextual elements. Observed deviations from these expectations create prompts for further discussion rather than direct evidence of design failure. Second, the framework provides a transferable conceptual model for future eA&F development. Designers must reason prospectively about how each information element will be visually processed and must reserve Scrutinise-level cognitive demand only for elements most critical to behaviour change, to minimise unnecessary cognitive load. In this way, PREVIEW-A&F methodology encourages deliberate alignment between interface design and intended implementation strategy mechanisms.

Table 1 presents a sample of the a priori AOI framework for the iNAAN report. Orientation-only AOIs are included because they allow testing of whether clinicians correctly distinguish items requiring only orientation from items requiring deeper processing. The complete 29-item AOI framework is provided in Supplementary File 1.

AOI	Description	Intended Attention	BCTs	Expected Clinician Action
1	Report title	Glance	n/a	None; orientation only
2	Location, date	Look	n/a	None; orientation only
7	Protocol adherence with temporal comparator	Scrutinise	Discrepancy from Goal Action Planning Feedback on Behaviour	Self-reflect on protocol compliance; review clinical approach; consider corrective action.

Table 1 Sample of the a priori Area of Interest (AOI) framework for the iNAAN Health Service Report, showing one example of each intended attention level (Glance, Look, Scrutinise). BCT = behaviour change technique. The complete 29-item framework is provided in Supplementary File 1.

Apparatus and Gaze Data Collection

Participants sat approximately 70 cm in front of one of two 24-inch monitors (Hewlett Packard HP E24 G5, set to either 1920 × 1080 or 1920 x 1200 resolution). Gaze data were recorded using a spectacle-mounted wearable eye tracker (Tobii Pro Glasses 3, Tobii AB, Sweden). A key methodological feature is the portability of this device: unlike screen-based eye trackers that require participants to remain in front of a fixed display, the Tobii Pro Glasses 3 allows natural head movement, making it suitable for use in settings that approximate real-world clinical environments. A spherical refractive correction was incorporated into the device, estimated by a registered optometrist (AD) from measurement of each participant's spectacle prescription, with near correction adjusted for age. Binocular visual acuity was confirmed in the plane of the monitor prior to data collection, with one participant's gaze data (P09 in Table 2) being excluded because the spherical lenses could not correct their significant astigmatism, which would have prevented them from clearly resolving the report content during the eye tracking session. Full technical specifications, including fixation filter parameters and calibration protocol, are provided in Supplementary File 2.

Procedure

Participants first completed a practice task (using a non-iNAAN document) to confirm their ability to scroll a two-page text PDF using a mouse scroll wheel. The eye tracker was then calibrated. Participants were then instructed to look away from the screen while their iNAAN

site-specific report was opened and displayed full screen. Participants were instructed to scroll the report using the scroll wheel only, not to resize the window, and to inform the examiner when they had examined the report to their satisfaction. No time limit was imposed. Full written instructions are provided in Supplementary File 3.

Think-Aloud Protocol

To capture the cognitive reasoning underlying gaze patterns, participants completed a think-aloud protocol immediately after the eye-tracking described above. They were instructed to provide a continuous verbal commentary of their thoughts, decisions, and any points of confusion or interest as they reviewed the report. Sessions were audio-recorded and transcribed verbatim.

Data Analysis

Gaze Data Processing

Gaze data were processed using Tobii Pro Studio and custom software to map gaze to regions of interest within dynamically scrolled documents (as shown in Figure 3 and described in Supplementary File 2). AOIs were drawn tightly around each item of interest (see Figure 2). Fixations were identified using a standard algorithm, allocating each fixation to the AOI containing the highest number of gaze pixels. Fixations with equal pixel counts across two AOIs were resolved by spatial proximity to the AOI centroid. Fixations not overlapping any AOI were assigned to background.

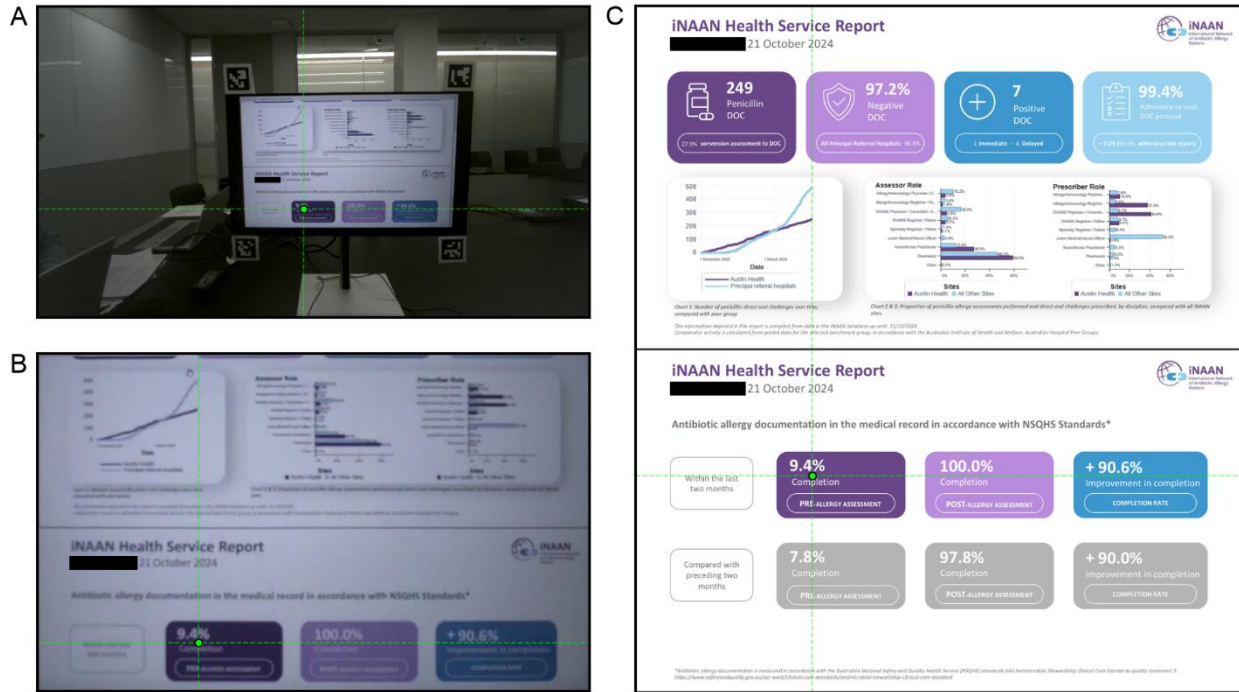


Figure 3 Mapping fixation positions from the head-mounted camera to the document. Panel A shows a frame from the head-mounted camera of a participant completing the task, with the overlay showing the estimated fixation position. Panel B shows the view of the document extracted from the head-mounted camera frame, obtained from applying the best transformation (across the potential document scroll positions) to the head-mounted camera frame and fixation position. Panel C shows the fixation position as transformed from the document view into the space of the complete document, from which the AOI can then be obtained (see Figure 2).

Objective Gaze Metrics

The primary gaze metric was total fixation duration per AOI: cumulative time (seconds) the eye was relatively stable within each AOI, summed per participant. Fixation duration was analysed differently by intended attention level: for Glance- and Look-classified AOIs, raw fixation duration was used; for Scrutinise-classified AOIs, fixation duration was normalised by the number of distinct information elements within each AOI, yielding a duration-per-information-unit metric. Inter-rater agreement for information-unit counts was assessed across four raters (EM, AA, AD, JT) using proportional agreement; initial agreement was 76.4% within ± 1 unit, acceptable for this estimation task, with consensus then reached on all discrepant items. The Glance/Look/Scrutinise classifications were determined independently by three raters (EM, JT, MK), achieving 100% agreement across all 29 AOIs.

Qualitative Coding of Think-Aloud Data

Think-aloud transcript were coded by two researchers (MK and VB) using a framework approach comprising four stages: familiarisation with transcripts, initial coding, framework

application, and interpretation. Two complementary analytical approaches were used. First, deductive coding applied the Theoretical Domains Framework (TDF), a theory-informed synthesis of 33 behaviour change theories comprising 14 domains, to identify which theoretical domains were activated as participants engaged with BCT-linked interface elements. Second, inductive coding allowed additional themes to emerge from the data without a pre-specified framework. Usability and interpretability themes identified inductively were analysed separately from deductive TDF domains and were not treated as theoretical constructs within the TDF. All data were double coded independently by both researchers. Discrepancies were first addressed through consensus discussion between MK and VB; unresolved discrepancies were adjudicated by a third researcher (LG).

Mixed-Methods Integration: Joint Display and Analysis

Gaze and think-aloud data were interpreted using a convergent integration approach. For each AOI, fixation duration metrics were grouped by a priori intended attention level and plotted for visual inspection (Figure 4). A planned contrasts analysis compared each AOI against the grand mean for its attention level group using a mixed model with random intercepts for participants and restricted maximum likelihood fitting (R version 4.3.1; lmer and emmeans packages), with $\alpha = 0.05$ and no correction for multiple comparisons given planned analyses. Integration then examined alignment between observed fixation duration and intended attention level, and between think-aloud processing depth and a priori expectations, with discrepancies between gaze and verbal data representing the primary analytic interest.

Results

Participant characteristics: Fourteen participants were tested, with 12 providing acceptable gaze data (86% success rate). Table 2 presents participant characteristics.

Participant ID	State / Country	Discipline	iNAAN Investigator	Report-Naïve	Eye-tracking data used
P01	NSW, Aus	ID physician	Yes	No	Yes
P02	QLD, Aus	ID physician	Yes	Yes	Yes
P05	VIC, Aus	Nurse	No	No	Yes
P06	WA, Aus	Pharmacist	Yes	No	Yes
P07	NSW, Aus	Nurse	Yes	No	Yes
P09	QLD, Aus	Nurse	Yes	Yes	No
P10	NSW, Aus	Pharmacist	Yes	No	Yes
P12	VIC, Aus	Pharmacist	Yes	No	Yes
P13	VIC, Aus	ID physician	Yes	No	Yes
P14	VIC, Aus	ID/Allergy registrar	Yes	Yes	Yes
P15	VIC, Aus	Pharmacist	Yes	Yes	Yes
P16	VIC, Aus	Allergist/Immunologist	Yes	No	Yes
P17	VIC, Aus	Pharmacist	Yes	No	No
P18	VIC, Aus	Allergist	Yes	Unknown	Yes

Table 2 Participant characteristics. ID=infectious diseases; iNAAN = International Network of Antibiotic Allergy Nations. Missing participant numbers P04, P08, and P11 were scheduled for testing, but did not attend. P17's eye tracking (ET) data were excluded due to poor quality calibration and multiple dropped samples. Participant P09's ET data were excluded due to significant astigmatism that could not be corrected by the spherical lenses during the gaze tracking component of the study. P03 was an international investigator excluded from participation as the study was limited to Australian sites.

Visual engagement patterns: Engagement patterns are shown in Figure 4. As expected, fixation durations increase when moving from AOIs classified as Glance, through Look, to Scrutinise. Raw values (upper panel) vary substantially (11.8-fold) within the Scrutinise group. Normalising values by the estimated number of pieces of information in each AOI (lower panel) reduces this variability (to 2.7-fold). Aside from noting unusual behaviour from simple visual inspection, the contrasts analysis highlights (asterisks) those AOIs whose mean differs significantly from the group mean. For example, AOI 18 received less fixation time than those other AOIs classified as Look (top panel), and AOI 8 received significantly more fixation time that might be expected from the pieces of information it contains (bottom panel). Such behaviours suggest a possible divergence between intended and observed engagement (unexpectedly low engagement with AOI 18, and unexpectedly high for AOI 8) and so create a prompt to review the corresponding report element. Whilst this prompt may arise from the visual engagement patterns alone, think-aloud results, given below, might further inform the nature of this review (for example, reports of AOI 8 being confusing). Unnormalised fixation durations were highly correlated to the number of fixations made into or within an AOI (Pearson $r = 0.985$; 95% CI 0.968 – 0.993; $p < 0.0001$), suggesting that analysing fixation counts would not add important information.

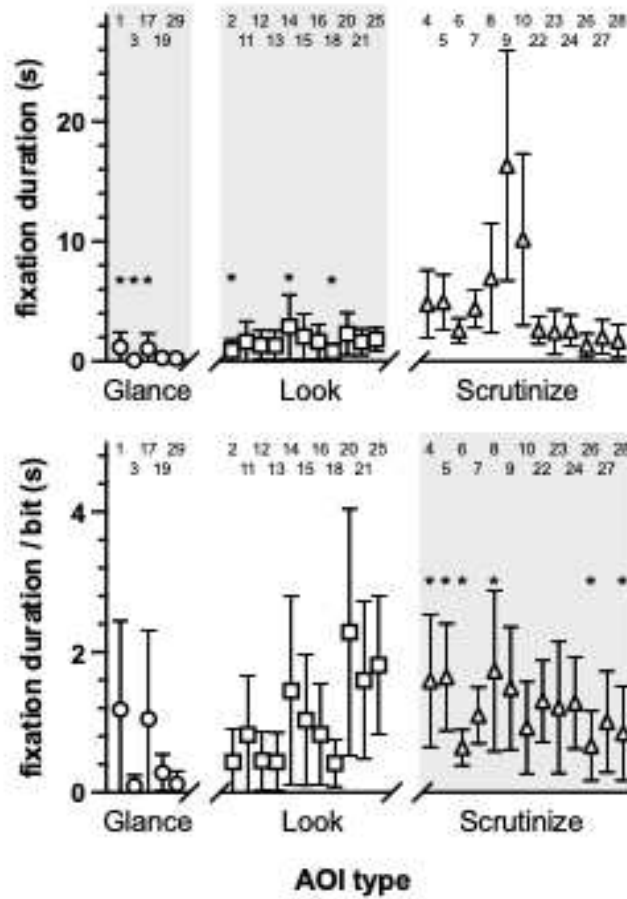


Figure 4: Total of fixation durations as a function of AOI type. Values are represented as unnormalised (upper panel) and normalised by the number of pieces of information (lower panel) in each AOI. Data within a grey panel had the mean value from each AOI compared to the group mean using a mixed effects model with planned contrasts, with asterisks representing significantly deviating values ($p \leq 0.05$). Symbols represent means, with error bars giving standard deviations. Individual AOI numbers are given along the top of each panel.

Think-aloud results

Think-aloud data were obtained from all 14 participants across 29 AOIs. Verbalisations were coded using a two-dimensional framework comprising processing level (orienting, extracting, interpreting, problem-solving) and processing quality (successful, mixed, confused, dismissed) (Supplementary File 4). Where no verbalisation was produced for an AOI, this was recorded as “No Think-aloud Evidence” rather than inferred as engagement absence. Across the 29 AOIs, the number of verbalisers ranged from 0 (AOIs 16, 17, 18, 19, 25, 29) to 14 (AOI 4). Orientation elements (Glance-classified AOIs) predictably generated few or no verbalisations, consistent

with their design intention. BCT-linked Scrutinise elements attracted the highest verbalisation rates, with 8 of 13 Scrutinise AOIs generating verbalisations from six or more participants.

Processing level varied systematically across AOI types. Scrutinise elements generated the highest proportions of Interpreting and Problem-solving codes, with participants frequently connecting data to clinical context, making cross-site comparisons, or identifying performance discrepancies. For example, participants reviewing the prescriber discipline chart (AOI 10) produced verbalisations such as "the use of nurse practitioners is probably the way forward but we need to work out how we can implement that safely" and "summarising who's assessing and who's prescribing internally, particularly if I'm looking at expanding out our workforce," demonstrating activation of BCTs relating to demonstration of behaviour and social comparison. Look-classified elements predominantly generated Extracting codes, with participants reading and confirming values without deeper interpretation, consistent with the orientation-level engagement intended for these elements.

TDF coding identified domains that were consistently activated across participants engaging with Scrutinise-classified AOIs. Knowledge of consequences and Beliefs about consequences were the most frequently coded domains, reflecting participants' engagement with performance discrepancy data and peer comparisons. Social influences was coded prominently at AOIs featuring peer-benchmarking data (AOIs 4, 5, 10), consistent with the social comparison BCTs embedded in those elements. Behavioural regulation was coded where participants articulated intentions to act in response to performance data. Memory, attention and decision processes was coded at AOIs where participants described difficulty interpreting metrics (AOIs 7, 14, 27), providing theoretical grounding for the Hidden Opportunity and Divergent classifications. Look-classified and Glance-classified AOIs generated minimal TDF-coded verbalisations, consistent with their design as orientation rather than BCT-activation elements.

Integrated findings:

Integration of fixation duration and think-aloud data was organised around four integration patterns, derived through collaborative review by the full multidisciplinary authorship team (Supplementary File 4). Convergent: eye-tracking fixation duration is consistent with the intended attention level and think-aloud confirms processing at the expected depth i.e., the BCT activation pathway is intact. Hidden Opportunity: eye-tracking fixation duration is consistent

with the intended attention level, but think-aloud processing does not reach the expected depth i.e., the element is visually attended to as intended but BCT activation is absent or incomplete. Attentional Mismatch: eye-tracking fixation duration does not match the intended attention level but think-aloud confirms processing at the expected depth among those who verbalised i.e., the discrepancy exists in the fixation data alone, not in cognitive processing. Divergent: the a priori framework matches neither eye-tracking fixation duration nor think-aloud processing depth i.e., engagement failure is present across both data streams, whether through under-dwell with absent BCT activation, over-dwell with confused or dismissed processing, or cohort-level failure of both data streams to confirm intended engagement. A fifth pattern, Silent Engagement, was applied to orientation elements (Glance/Look) where no verbalisation is expected; as absence of verbalisation requires no redesign response, these AOIs are not discussed further. Of the 29 AOIs, 14 were classified as Convergent, one as Hidden Opportunity, eight as Attentional Mismatch, three as Divergent, and three as Silent Engagement. These patterns are pragmatic analytic heuristics to guide prospective redesign, not an empirically validated taxonomy.

Convergent

Fourteen AOIs were classified as Convergent, in which fixation duration matched the intended attention level and think-aloud confirmed processing at the expected depth. This included the majority of Look-classified contextual elements and the majority of Scrutinise elements, where participants produced Interpreting and Problem-solving verbalisations aligned with the intended BCTs.

Hidden Opportunity

One AOI was classified as Hidden Opportunity. AOI 27 (pre-assessment documentation accuracy, comparator period) was visually attended within the expected Scrutinise range, indicating that fixation duration matched the a priori framework. However, the single participant who verbalised reported confusion about the metric's operational meaning and think-aloud evidence was insufficient to confirm BCT activation across the cohort. This pattern, adequate visual attention with compromised cognitive processing, represents a comprehension barrier that eye-tracking alone would not detect. Redesign of the metric label to clarify its operational definition is indicated before broader implementation.

Attentional Mismatch

Eight AOIs were classified as Attentional Mismatch, in which fixation duration deviated from the intended level while think-aloud data indicated processing that was successful at the depth engaged. Three of these, AOIs 1 (report title, page 1), 3 (iNAAN logo), and 17 (report title, page 2), are orientation-only elements carrying no BCT linkage. AOIs 1 and 17 showed over-dwell relative to their respective Glance group means, and AOI 3 showed under-dwell; in each case, think-aloud confirmed successful orienting-level processing. The mismatch between observed and intended fixation duration is noted, however the absence of BCT linkage means none of these elements warrants redesign. AOI 2 (location, date) similarly showed under-dwell relative to the Look group mean with successful orienting-level processing from the six participants who verbalised; as a Look-classified orientation element with no BCT linkage, this mismatch is also low priority for redesign.

The remaining Attentional Mismatch AOIs involve BCT-linked Scrutinise elements and carry higher redesign relevance. AOI 5 (negative DOC rate) showed normalised over-dwell relative to the Scrutinise group mean, with ten participants verbalising engagement; however, processing remained at the Extracting and Orienting level across all verbalisers, with none progressing to the Interpreting or Problem-solving expected for a Scrutinise element. The over-dwell therefore reflects processing difficulty rather than meaningful engagement with the BCT activation pathway, and the visual presentation of this element warrants redesign to reduce cognitive load and support progression to social comparison activation. AOI 28 (post-assessment documentation accuracy, comparator period) showed normalised under-dwell relative to the Scrutinise group mean; however, think-aloud data confirmed Interpreting-level processing among ten verbalisers despite the lower fixation duration, with mixed quality across those responses. The Attentional Mismatch here reflects efficient rather than deficient processing: the element is being extracted rapidly with BCT activation present, though the mixed quality of processing warrants monitoring to confirm this efficiency is sustained as the report matures.

Of note, AOIs 4 (DOC numbers and conversion statistics) and 8 (DOC trend over time) each showed normalised over-dwell relative to the Scrutinise group mean; however, think-aloud data confirmed this reflected genuine depth of engagement rather than processing difficulty, with all 14 participants verbalising active self-comparison for AOI 4. The Attentional Mismatch classification is retained for consistency however, both elements carry low redesign priority: the

over-dwell reflects the genuine cognitive demands of comparative, multi-element displays rather than any failure of the BCT activation pathway.

Divergent

Three AOIs were classified as Divergent, in which the a priori framework matched neither eye-tracking fixation duration nor think-aloud processing depth, with engagement failure present across both data streams. AOI 6 (positive DOC rate and adverse events) showed normalised under-dwell combined with think-aloud processing remaining at the Extracting and Orienting level, with no BCT activation despite Scrutinise classification. Ten participants verbalised engagement, however none progressed beyond surface-level confirmation of a zero or near-zero adverse event rate. The a priori framework matches neither the eye-tracking nor the think-aloud data: the element is neither attended to at the intended depth nor cognitively processed in a manner consistent with its intended BCT pathway. This pattern reflects a rational ceiling effect i.e., when adverse event rates are at or near zero, the goal-setting and self-monitoring BCTs embedded in this element have no performance discrepancy on which to operate. However, the element's design does not currently support BCT activation when data are non-zero, and redesign may be required to address this. AOI 14 (Chart 1 title/description) showed over-dwell relative to the Look group mean. The single participant who verbalised reported confusion about the element's meaning and explicitly dismissed it as unhelpful despite the additional fixation time. AOI 26 (pre-assessment documentation present, comparator period) showed normalised under-dwell combined with only three verbalisers, insufficient to confirm BCT activation across the cohort. Among those who verbalised, processing reached the Interpreting level, however the cohort-level failure of both visual engagement and cognitive processing to meet the a priori framework means this element cannot be considered functioning as intended. AOIs 6, 14, and 26 represent priority redesign targets; their engagement barriers are either entirely invisible to eye-tracking data alone or insufficiently distributed across the cohort to support BCT activation at scale, illustrating why integration of the two data streams is essential for detecting comprehension failures of this type.

Discussion

This study presents a theory-informed methodology for prospectively evaluating whether clinicians engage with key elements of an eA&F intervention in a manner consistent with its

intended behavioural design. By integrating eye-tracking with retrospective think-aloud data, the approach links objective measures of overt visual attention with qualitative evidence of cognitive processing, providing a structured mechanism for identifying redesign opportunities before broader implementation.

We propose that implementation strategies should be conceptualised as optimisable behavioural systems rather than static interventions that are only refined after deployment. From this perspective, the principal implementation science contribution of this work is not the use of eye-tracking technology itself, but the prospective optimisation of an implementation strategy. eA&F interventions are commonly developed using expert input and user consultation before progressing directly to implementation or effectiveness testing. Engagement barriers may therefore remain undetected until the strategy has already been deployed, at which point modifications require additional redesign, redeployment, and re-evaluation. The methodology described here seeks to shift optimisation activities earlier in the implementation pathway by identifying interface-level barriers during intervention development rather than after implementation has commenced.

From a theoretical perspective, the methodology distinguishes between four related but conceptually distinct constructs: overt visual attention, cognitive processing, activation of intended behavioural mechanisms, and downstream implementation outcomes. The present study directly measured only the first two. Behaviour change technique activation and implementation consequences remain inferred, based on the degree to which observed engagement was consistent with the intervention's intended behavioural design. This distinction is important because eye-tracking should not be interpreted as a direct measure of cognition or behaviour change, but rather as an objective indicator of where attentional resources are allocated.

The findings also contribute to the emerging discussion regarding implementation pace. There is increasing recognition that efforts to improve implementation outcomes should also consider the efficiency with which interventions and implementation strategies move from development to routine use. Although the present proof-of-concept study did not directly evaluate implementation speed or redesign burden, it illustrates one plausible mechanism through which implementation pace might be improved. By identifying and resolving eA&F engagement barriers before broader deployment, prospective optimisation approaches may reduce the need

for avoidable post-implementation redesign cycles and facilitate more efficient progression towards implementation readiness.

Limitations

Several limitations should be noted. First, this study represents a proof-of-concept application within a single implementation programme; further evaluation is required to establish transferability to other implementation strategies, user groups, and digital intervention formats.

Second, the iNAAN eA&F report had already undergone iterative refinement with stakeholder input before this study was undertaken, such that the methodology was applied concurrently with ongoing optimisation rather than being embedded from the start of intervention development.

Third, eye-tracking provides objective evidence of overt visual attention but does not capture covert peripheral processing; however, the fine-grained quantitative detail of eA&F reports makes peripheral processing of critical content unlikely. Gaze metrics should therefore be interpreted as observable proxies of attentional allocation rather than direct measures of cognition or behaviour change.

Fourth, the intended attention levels specified for each AOI represent a priori judgements by the multidisciplinary team and should be considered design hypotheses rather than validated cognitive models; a principal function of this methodology is to test and refine these assumptions empirically.

Conclusions

This study describes a theory-informed methodology for prospectively evaluating engagement with electronic audit and feedback interventions. By linking interface elements to their intended behaviour change functions and integrating objective measures of overt visual attention with qualitative think-aloud data, the approach provides a structured means of identifying redesign opportunities before broader implementation. For implementation scientists and health service researchers developing eA&F tools, the five-phase framework provides a transferable structure for embedding prospective engagement evaluation within intervention development cycles. For digital health developers and clinical programme leaders, the findings highlight the cost of omitting this step: engagement barriers invisible to self-report methods alone will persist

undetected into implementation, where they become substantially more costly to identify and resolve. Further research is warranted to establish whether earlier application of this methodology across diverse clinical and implementation contexts reduces downstream redesign burden and, in doing so, supports more efficient pathways to sustained implementation.

List of Abbreviations

A&F: Audit and feedback; AOI: Area of interest; BCT: Behaviour change technique; DOC: direct oral challenge; eA&F: Electronic audit and feedback; iNAAN: international Network of Antibiotic Allergy Nations; TDF: Theoretical Domains Framework.

Declarations

Ethics approval and consent to participate

This study was conducted in accordance with the National Statement on Ethical Conduct in Human Research (National Health and Medical Research Council, Australia) and the Declaration of Helsinki. It was conducted as a sub-study (protocol 78719) of the iNAAN project (HREC/71789/Austin-2021), approved by the Austin Health Human Research Ethics Committee. All participants provided informed electronic consent prior to participation.

Consent for publication

Not applicable. The manuscript does not contain individually identifiable data.

Availability of data and materials

The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

No external funding was received for this pilot study. Funding for the larger programme of research has been sought through an NHMRC Ideas Grant (Application ID: 2048067).

Authors' contributions

MK led study conception, design, qualitative data collection & analysis and manuscript preparation.

AA led study conception, design, eye-tracking data collection & analysis and manuscript preparation.

AD led study conception, design, eye-tracking data collection & analysis and manuscript preparation.

DM led gaze data processing software development & analysis.

CP led qualitative data collection.

VB & LG contributed to qualitative coding and data adjudication.

EM & JT contributed to study conception.

JJ contributed to study conception.

All authors read and approved the final manuscript.

Acknowledgements

We thank all iNAAN investigators who participated in the study and the Drug Hypersensitivity Symposium organising committee for facilitating recruitment.

Supplementary Files

- Supplementary File 1: Complete 29-item AOI rationale table for number of pieces of information, expected gaze behaviour (Glance, Look, Scrutinise) and desired behaviour change.
- Supplementary File 2: Full technical apparatus specifications (Tobii Pro Glasses 3 setup, calibration protocol, software description, fixation filter parameters)
- Supplementary File 3: Full participant instruction script
- Supplementary File 4: Data Integration Table

References

1. Ivers N, Yogasingam S, Lacroix M, Brown KA, Antony J, Soobiah C, et al. Audit and feedback: effects on professional practice. *Cochrane Database of Systematic Reviews*. 2025(3).
2. Crawshaw J, Meyer C, Antonopoulou V, Antony J, Grimshaw JM, Ivers N, et al. Identifying behaviour change techniques in 287 randomized controlled trials of audit and feedback interventions targeting practice change among healthcare professionals. *Implementation Science*. 2023;18(1):63.
3. McCleary N, Desveaux L, Presseau J, Reis C, Wittman HO, Taljaard M, et al. Engagement is a necessary condition to test audit and feedback design features: results of a pragmatic, factorial, cluster-randomized trial with an embedded process evaluation. *Implementation Science*. 2023;18(1):13.
4. Sarkies M, Francis-Auton E, Long J, Roberts N, Westbrook J, Levesque J-F, et al. Audit and feedback to reduce unwarranted clinical variation at scale: a realist study of implementation strategy mechanisms. *Implementation Science*. 2023;18(1):71.
5. Shuldiner J, Lacroix M, Saragosa M, Reis C, Schwartz KL, Gushue S, et al. Process evaluation of two large randomized controlled trials to understand factors influencing family physicians' use of antibiotic audit and feedback reports. *Implementation Science*. 2024;19(1):65.
6. Tuti T, Nzinga J, Njoroge M, Brown B, Peek N, English M, et al. A systematic review of electronic audit and feedback: intervention effectiveness and use of behaviour change theory. *Implement Sci*. 2017;12(1):61.
7. Soresi J, Bertilone C, Banks E, Marshall T, Murray K, Preen DB. Features and effectiveness of electronic audit and feedback for patient safety and quality of care in hospitals: A systematic review. *Health Informatics Journal*. 2025;31(1):14604582251315414.
8. Tsang JY, Peek N, Buchan I, van der Veer SN, Brown B. Systematic review and narrative synthesis of computerized audit and feedback systems in healthcare. *J Am Med Inform Assoc*. 2022;29(6):1106-19.
9. Ryskina K, Jessica Dine C, Gitelman Y, Leri D, Patel M, Kurtzman G, et al. Effect of Social Comparison Feedback on Laboratory Test Ordering for Hospitalized Patients: A Randomized Controlled Trial. *J Gen Intern Med*. 2018;33(10):1639-45.
10. Navathe AS, Volpp KG, Bond AM, Linn KA, Caldarella KL, Troxel AB, et al. Assessing The Effectiveness Of Peer Comparisons As A Way To Improve Health Care Quality. *Health Aff (Millwood)*. 2020;39(5):852-61.
11. Billue KL, Safford MM, Salanitro AH, Houston TK, Curry W, Kim Y, et al. Medication intensification in diabetes in rural primary care: a cluster-randomised effectiveness trial. *BMJ Open*. 2012;2(5).
12. Roos-Blom MJ, Gude WT, de Jonge E, Spijkstra JJ, van der Veer SN, Peek N, et al. Impact of audit and feedback with action implementation toolbox on improving ICU pain management: cluster-randomised controlled trial. *BMJ Qual Saf*. 2019;28(12):1007-15.
13. Chima S, Martinez-Gutierrez J, Hunter B, Manski-Nankervis J-A, Emery J. Optimization of a Quality Improvement Tool for Cancer Diagnosis in Primary Care: Qualitative Study. *JMIR Form Res*. 2022;6(8):e39277.
14. Driessen SR, Van Zwet EW, Haazebroek P, Sandberg EM, Blikkendaal MD, Twijnstra AR, et al. A dynamic quality assessment tool for laparoscopic hysterectomy to measure surgical outcomes. *Am J Obstet Gynecol*. 2016;215(6):754.e1-e8.
15. Grissom N, Bhatnagar S. Habituation to repeated stress: get used to it. *Neurobiol Learn Mem*. 2009;92(2):215-24.
16. Larson KE, Bradshaw CP. Cultural competence and social desirability among practitioners: A systematic review of the literature. *Children and Youth Services Review*. 2017;76:100-11.
17. van de Mortel TF. Faking It: Social Desirability Response Bias in Self-report Research. *The Australian Journal of Advanced Nursing*. 2008;25(4):40-8.
18. Chaichana J, Eley R, Watling C, Ng L. Evaluating Eye Tracking Technology in Nursing Education: A Scoping Review on Medication Administration Training. *Nurs Rep*. 2025;15(6).
19. Chidda A, Seidel A. Exploring eye-tracking in orthopedics: Potential uses and perspectives. *Journal of Musculoskeletal Surgery and Research*. 9.
20. Wu C, Cha J, Sulek J, Zhou T, Sundaram CP, Wachs J, et al. Eye-Tracking Metrics Predict Perceived Workload in Robotic Surgical Skills Training. *Hum Factors*. 2020;62(8):1365-86.
21. Atkins L, Francis J, Islam R, O'Connor D, Patey A, Ivers N, et al. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. *Implement Sci*. 2017;12(1):77.

22. Mitri EA, Fletcher LR, Vogrin S, Barnes S, Powell N, Peter J, et al. Direct oral challenge for penicillin allergy: the International Network of Antibiotic Allergy Nations (iNAAN) study. *Clinical Infectious Diseases*. 2026:ciag082.
23. Mitri EA, Vogrin S, Paynter C, Fletcher LR, Peter J, Powell N, et al. Direct Oral Challenge for Penicillin Allergy: A Hospital Implementation Evaluation Via the International Network of Antibiotic Allergy Nations (iNAAN) Study. *Clinical Infectious Diseases*. 2026:ciag083.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFile2TechnicalSpecifications.docx](#)
- [SupplementaryFile3Procedure.docx](#)
- [SupplementaryFile1AOIFramework.xlsx](#)
- [SupplementaryFile4IntegratedDataMatrix.xlsx](#)