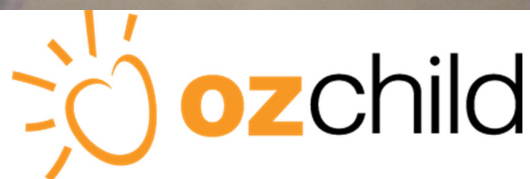


Final Report: Examining the Impact of Evidence-Based Decision-Making Professional Development for Human Services Leaders in the Child Welfare Sector.



Acknowledgements

This research was sponsored by the Centre for Organisational Change & Agility (COCA), Torrens University Australia (TUA) in partnership with OzChild, the Center for Evidence-Based Management and Carnegie Mellon University. The project collaborated with industry partners and was coordinated through the Project Advisory Committee (PAC) made up of TUA researchers and representatives from OzChild, Carnegie Mellon University and the Center for Evidence-Based Management. The PAC met regularly during the project to co-design the research, create the research ethics application and approval and during the stages of data collection. We would like to thank the Centre for Evidence-Based Management for providing course materials, scenarios, assessments and teaching of the professional development course for participants. We would like to especially thank OzChild for course co-ordination and for facilitating participation in this research.

We would like to express our deep gratitude to the participants of this study in giving their time and insights for this research.

Members of the Project Advisory Committee

Torrens University Australia	OzChild	Center for Evidence- Based Management	Carnegie Mellon University
Professor Roslyn Cameron: Director of COCA Dr Fiona Chatteur Adjunct Senior Research Associate COCA	Dr Lisa J. Griffiths, CEO Jarrod Edwards Exec Director People & Performance Tim Lippis Learning, Development & Inclusion Officer	Professor Eric Barends Managing Director	Professor Denise M. Rousseau Professor of OB & Public Policy

Resourcing

This research was resourced by Torrens University Australia's Centre for Organisational Change & Agility (COCA) and OzChild Australia.

Citation

Chatteur, F. M., Cameron, R., Griffiths, L. J., Barends, E., Lippis, T., Rousseau, D. M., Edwards, J. (2026) *Final Report: Examining the Impact of Evidence-Based Decision-Making in the Child Welfare Sector*. Torrens University Australia. Research Report
<https://doi.org/10.25905/31957749>

Contents

Acknowledgements	2
Resourcing	2
Executive Summary and Recommendations	6
Aim:	6
Research Questions (RQ)	6
Key Findings	7
Recommendation 1: Access to Scientific Research & Community of Practice (CoP)	8
Recommendation 2: Documentation of Evidence	9
Recommendation 3: Post-Training Checklists	9
Recommendation 4: Organisation-Wide Training	10
Recommendation 5: Future Research Child Welfare Sector	10
1. Introduction	11
1.1 Australian Child Welfare Context	12
Evidence-Based Programs and Practices	12
2. Literature Review	13
3 Evidence-Based Decision-Making	14
3.1 Theoretical Lens	15
4. Research Approach and its Rationale	17
4.1 Research Design	18
4.2 Ethics Clearance	21
5. Analysis and Outcomes	21
Participants	21
Inter Coder Reliability	23
Protocol Analysis of the "think aloud" data	24
Statistical Analysis of the Cohort Data	34
Statistical Analysis of Larger Course Population	37
Data Visualisations	40
6. Key Findings	73
7. Limitations	75
8. Conclusions	76
9. Author Contributions:	77
10. References	77
11. Appendices	82

11-1 Ethics Approval	82
11.2 Pre- and Post-tests	85
11.3 "Think aloud" scenarios	97
11.4 Keyword Lists	98
11.5 Demographic Survey	101
11.6 Krippendorff Alpha Calculations (Technical notes)	103
11.7 Pre-Post Test Results by Learning Outcome	107

Executive Summary and Recommendations

This study investigates how evidence-based decision-making (EBDM) professional development training influences practice for human service leaders in the child welfare sector in Australia.

Aim:

The overall research aim of this research was to explore the differences in thinking of senior leaders when applying their EBDM training to professional practice.

Research Objectives (RO):

RO1: Compare the changes in decision-making process quality between the intervention group (professional program participants) and the Control group, highlighting the program's impact and effectiveness on professional practice.

RO2: Identify specific changes in decision-making behaviours, strategies and practices that occur as a result of participating in the program.

RO3: Provide insights into the factors that facilitate or hinder the implementation of evidence-based decision-making in the child welfare sector.

RO4: Identify mediating factors or mechanisms that contribute to the observed changes in decision making after completing the program.

RO5: Generate recommendations for refining or adapting the EBDM program to further enhance its impact and effectiveness in improving decision-making processes and outcomes in the child welfare sector.

To achieve these research objectives the following research questions are posited:

Research Questions (RQ)

RQ1: To what extent does the Certificate in Evidence-Based Decision-Making for Human Service Leaders produce changes in the application of evidence-based principles and practices in decision-making among executive leaders in the child welfare sector?

RQ1a: What are the specific changes in decision-making behaviours and strategies observed among leaders in the child welfare sector following the completion of the Evidence-Based Decision-Making for Human Services Leaders program?

RQ1b: What are the differences, if any, in the change in decision-making quality between the leaders who participated in the Certificate in Evidence-Based Decision-Making for Human Services Leaders program and a control group of leaders who did not receive the program?

Research Design

The challenge in this investigation was in determining the differences in thinking by senior leaders who undertook an evidence-based decision making (EBDM) professional development course when compared to senior leaders who did not. This necessitated the use of two groups,

an intervention group and a control group. The study therefore used a quasi-experimental approach and collected both qualitative and quantitative data.

The study employed a quasi-experimental mixed methods research design which collected quantitative and qualitative data across three research stages. Stage 1 involved a pre-intervention "think aloud" session for both groups using a unique scenario which presented situations likely to occur in professional practice. Stage two involved pre- and post-tests prior to the course cohort undertaking their training and afterwards. Both the intervention group and the control undertook both sets of tests. Stage three involved a post-course "think aloud" session using a different scenario, also based on professional practice, again undertaken by both the intervention group and the control.

Key Findings

- There was no significant difference in the pre-intervention and in-course test scores between the OzChild participants and the larger EBDM course population. However, a significant difference (around 22%) between pre and in-course test scores was found for the overall EBDM course population. This indicates significant improvements in learning for those participants who undertook the course.
- When the larger EBDM course population is examined by student group, in terms of Undergraduate, Masters or Executive level participants, there is no statistical significance between the pre and post-test results. This appears to indicate that EBDM training is equally beneficial to participants notwithstanding educational or leadership levels.
- There was a statistically significant reduction on the *number of utterings* in the "think aloud" data for the EBDM course cohort, mostly due to one participant who 'drew a blank'.
- There were no statistically significant differences found for the Control group's "think aloud" data. However, the results indicate an increase in the use of organisational data, with slight reductions in the use of stakeholder and practitioner evidence.
- The Control group showed a reliance on organisational data and no reference to scientific data as sources of evidence. This continued reliance on organisational data, and the absence of scientific evidence post-intervention reinforce what we know from the literature: without a structured intervention, leaders default to experience and internal data.

- The EBDM course cohort showed small increases in considering stakeholders, practitioners and a slightly greater increase in using scientific data, but not enough to indicate statistical significance.
- The EBDM course group indicated more frequent utterings of EBDM phases, and pillars of evidence, evidenced in the “word clouds”. The Control group differed from the EBDM course group in that the differences between their pre- and post-intervention “think aloud” utterings seemed to centre on the variations in the scenarios rather than changes in thinking.
- Neither the EBDM group nor the Control group considered “Assessing” proposed solutions which is a key phase in EBDM. This represents a gap in the research process, where the application of the “Assessing” phase was not possible using the fictional scenarios. Nonetheless, the absence of the “Assessing” phase across both groups is particularly important. From both a research and practice perspective, this highlights a systemic gap in evaluation and learning of EBDM training that is critical in human services, where outcomes and accountability matter.
- The EBDM course group showed less “Applying” of solutions, appearing to want to wait until the data was appraised prior to forming an opinion on solutions. This is a key phase of EBDM and indicates more consideration by the EBDM course group of the evidence before proposing a solution. This was less evident in the pre-intervention “think aloud” sessions and in the Control group. This indicates that leaders can become more analytical, but without clear structures, this can lead to analysis without translation.
- There is evidence of a large reduction in post-test scores due to the Forgetting Curve (Murre & Dros, 2015) for participants who completed the EBDM course.

Recommendation 1: Access to Scientific Research & Community of Practice (CoP)

To facilitate the “acquire” phase of EBDM and to enhance the use of peer-reviewed scientific data, which is one of the four pillars of evidence in EBDM (Open Learning Initiative, 2025), it is recommended that the OzChild organisation consider access to peer-reviewed journals in the field of child welfare. A repository of relevant Open Access Journals which is frequently

updated, advertised and accessed internally would be a valuable addition to the organisation and would represent a relatively small investment. It is also recommended that OzChild establish a EBDM Community of Practice for senior Leaders to provide a space to reflect, discuss and exchange ideas on their EBDM practices. Recent peer-reviewed research could be discussed in the proposed EBDM Community of Practice.

Recommendation 2: Documentation of Evidence

It is recommended that leaders in the child welfare system and those responsible for decision-making document evidence gathered to justify their decision-making. This aligns with organisational governance and accountability. In accordance with the EBDM four pillars of evidence, this should include organisational data, scientific data, stakeholder and practitioner evidence. The type of evidence gathered may include stakeholder and practitioner surveys, questionnaires, interviews, focus groups, as well as industry-based statistics, organisational data, government white papers and peer-reviewed research. EBDM principles state that three sources of evidence are required. Documentation of evidence affords confidence in the “aggregate” stage and allows for accurate application of the Bayes Rule, a key principle in EBDM (Open Learning Initiative, 2025). It also allows for review of the decision-making in the “assess” phase, which was largely not employed by participants in this study.

Documentation SOPs could be standardised by developing a EBDM Transfer of Learning System for the Australian Child Welfare Sector which would facilitate and formalise decision-making and afford organisational oversight.

Recommendation 3: Post-Training Checklists

The pre- and post-intervention test data with the EBDM course group indicates that Ebbinghaus’ Forgetting Curve may be an issue for senior leaders post training (Murre & Dros, 2015). The findings also indicate that senior leaders who have undertaken the EBDM course do not advance to the application of solutions at an early stage (the “Apply” phase). It is therefore recommended that a multi-stage checklist process for decision-making, the application of solutions and subsequent assessing the outcomes of the decision-making would assist in the review of decision-making.

An EBDM Transfer of Learning System for the Australian Child Welfare Sector could be developed in future research to embed checklists into the existing workflow and leadership routines without introducing an additional layer of paperwork. The checklists would therefore become a key artefact in a more comprehensive EBDM Transfer of Learning system.

A new area of enquiry in future research could also investigate where ethically informed artificial intelligence (AI) could be embedded into this system.

Recommendation 4: Organisation-Wide Training

This research indicates that those participants who did not undertake training did not appear to apply EBDM practices in their decision-making thinking, apart from using organisational data post-intervention, nor did thinking appear to change based on informal or social learning, when comparing the pre- and post-intervention data. There was a complete lack of consideration in the Control group for drawing upon peer-reviewed scientific data as evidence in the post-intervention “think aloud” data.

It is therefore recommended that Human Service leaders in charge of child welfare decisions undertake EBDM training. The “forgetting curve” indicates that training alone does not deliver sustained change, and this research appears to indicate that training is not being implemented in practice, given the lack of consideration of the “apply” and “assess” phases. There is the opportunity to incorporate ongoing reinforcement and refreshers of EBDM principles, leadership modelling of EBDM and the integration of EBDM into everyday decision-making and program design.

Developing an EBDM Transfer of Learning System for the Australian Child Welfare Sector would strengthen expectations around applying evidence in real decisions as well as embedding evaluation and “Assess” as a routine part of practice.

Recommendation 5: Future Research Child Welfare Sector

It is recommended that future research be conducted that can build on this research and that of Burgemeister et al., (2022) to explore the prevalence of EBDM training and practices across the Child Welfare Sector in Australia and examine the key challenges and enablers for Child Welfare Agencies to embed EBDM into their Human Services workforces through professional development and organisational practices.

A key outcome of this proposed research would be to design a Framework for an EBDM Transfer of Learning System for the Australian Child Welfare Sector to ensure EBDM training fully transfers into professional practices post EBDM training and professional development interventions.

Particularly:

- applying evidence in real-time decision contexts
- embedding the “Assess” phase as part of routine leadership practice
- reinforcing learning over time to counter the forgetting curve

The proposed EBDM Transfer of Learning System for the Australian Child Welfare Sector has the value of complementing formal training with practice-based reinforcement, application opportunities and organisational integration, so that EBDM becomes part of how decisions are made, rather than something leaders understand conceptually. This proposed EBDM Transfer of Learning System would be important for sector-wide implementation, not just within the OzChild organisation.

1. Introduction

This report describes a research study undertaken in 2025 which investigated senior leaders from the OzChild organisation who undertook the Certificate in Evidence-Based Decision-Making for Human Service Leaders and compared the differences in thinking with senior leaders who did not undertake the professional development, using a quasi-experimental mixed methods approach.

This report firstly provides in Section 1.1 the context of the research, which is the Australian Child Welfare sector. It provides background on evidence-based practices and investigates it in this context. This is followed by a literature review which investigates the case for evidence-based decision-making for senior leadership.

The theoretical lens which frames the study is that of evidence-based decision-making, and the next section provides detail on the principles underlying evidence-based decision-making, the six A's and the four pillars of evidence. The philosophical paradigm underpinning this research is that of social constructivism, which is pragmatic, learner centred and socially accountable.

Research design is detailed in Section 4.1, which describes a quasi-experimental mixed-methods approach. The research has been undertaken in three stages, a pre-intervention "think aloud" exercise, pre- and post-intervention tests and a post-intervention "think aloud" exercise. Quantitative from a larger course cohort is compared with the participants and the "think aloud" data is analysed using protocol analysis.

Section 5 details the analysis and outcomes, including protocol analysis, employing Krippendorff's alpha for cross-coder reliability, quantitative analysis, testing for statistical significance and qualitative results, including data visualisations and participant quotes. Section 6 outlines key findings, Section 7 limitations and Section 8 the conclusion, which discusses how this research answered the research questions.

1.1 Australian Child Welfare Context

Child welfare organisations in Australia operate in the community services sector (CSS¹) and can be not-for-profit organisations (NFPs) or private entities. They function to ensure the safety and wellbeing of children. This field needs urgent improvement in Australia, where the number of children removed from home to live in the Out-of-Home Care (OOHC) system is rapidly rising—doubling in the last 10 years. There are consistent poor outcomes for children associated with OOHC (The Department of Social Services, 2025). Indigenous children are over-represented and ten times more likely to be removed (Australian Institute of Family Studies, 2017; Lewis et al., 2017). Internationally, the HSO more broadly is undergoing a period of volatility, facing rapid changes in society, technology and the economy (Shin & McClomb, 1998; Smith & Phillips, 2016). Within this context, ensuring earlier and timely intervention supporting vulnerable children’s needs is essential to both prevent future complex problems and improve outcomes (Australian Institute of Family Studies, 2018); Shonkoff et al., 2012). Further, these empirical studies looking at organisations delivering services within child welfare have found improved outcomes for children when the organisations have engaged organisational climates (Aarons & Sawitzky, 2006; Glisson & Green, 2011; Glisson & Hemmelgarn, 1998)

Evidence-Based Programs and Practices

Evidence-Based programs (EBPs) and evidence-based practices² are a growing part of the current requirement for innovation in service programs and increasingly considered strategies for HSOs to weather turbulent fiscal and policy environments (Carnochan et al., 2017). EBPs are based upon an understanding of how programs and practices work on the ground to deliver outcomes. They involve gathering research evidence to inform decision making about interventions, programs and practice principles delivered by individual practitioners, teams and whole organisations (Plath, 2012; Social Work Policy Institute, 2018). An EBP is a dynamic process of critical reflection that involves a combination of well-researched interventions with the on-the-ground experience of both practitioners and service users, with solutions being evaluated for their effectiveness (Plath, 2012). Back in 1999, the US Surgeon General declared that all those suffering from mental health disorders should receive state-of-the-art treatments that are evidence-based (Goldman & Azrin, 2003). It follows that some 20 years later, child welfare is requiring the same (Department of Family and Community Services (FACS), 2016; Department of Health and Human Services, 2016). The literature on EBPs in child welfare is limited, but broader research indicates that leadership has links between both organisational outcomes and client outcomes (G. A. Aarons, 2006; Corrigan et al., 2000). Much of this is predicated on the success of implementation, and a systematic literature review

¹ Community services sector (CSS) and community services organisation (CSO) are terms used in Australia, while human services sector (HSS) is more commonly used in other countries. The term community-based organisation (CBO) also used in the USA.

² A similar term is evidence-informed practice (EIP).

conducted in 2021 focused on the role of executive leadership and its role in the implementation of EBPs (McCarthy & Griffiths, 2021).

The findings revealed the additional leadership skills required by those leading organisational innovation and/or the development, implementation and sustainment of EBPs. These include a detailed understanding and appreciation of EBPs, searching for innovations and potential influence, experimenting with new concepts and procedures, and studying emerging social and economic trends (Finn et al., 2016; Novins et al., 2013; Rocque et al., 2014; (Shin & McClomb, 1998) These involved expanding evidence-based decision making, judgement and proactive problem-solving skills identified as existing important aspects of leadership in NFP HSOs (Shaw et al., 2013) and demand particular abilities, training and development (Shaw et al. 2013, p. 319).

Shin and McClomb (1998) found that organisations whose top executives had management training experience, planned and implemented innovations more frequently than did organisations whose top executives lacked such experience. This is supported by the 'knowledge and understanding' aspect of Bish and Becker's (2016) additional meta-category of non-profit orientation. The primary objective of this category relates to enhancing evidence-based decision making by understanding the broader NFP context in which the business operates, communication methods and overall approach to management.

A 2022 study examined whether the geographic location of Communities for Children (CfC) service providers in Australia affected the implementation of evidence-based programs (EBP). The cross-sectional survey examined the differences in EBP knowledge and uptake depending on the urbanicity and state of a program. Participants from regional areas were more resistant to EBP targets than those from major urban areas, and experienced greater implementation challenges than those from major urban settings but conversely did report greater knowledge about EPBs. They demonstrated that location is an important and influential factor in the implementation of EBP in the Child Welfare sector in Australia (Burgemeister et al., 2022).

The systematic literature review (SLR) by McCarthy and Griffiths in 2021 led to the development of OzChild's internal Leadership program (called the Leadership PASS) (McCarthy & Griffiths, 2021) and sourcing an external Evidence-Based Management Program (adapted to be called Evidence-Based Decision Making for Human Services Leaders) (Open Learning Initiative, 2025). Both programs have been run with multiple cohorts. However, neither has been fully evaluated to understand the changes in decision making that occur following completion and if therefore the change improves decision making in relation to client outcomes.

2. Literature Review

This literature review discusses research into senior leadership in the child welfare sector in Australia and abroad, to understand the challenges in gathering data on leadership. It will also investigate evidence-based decision making and its increased use in management and leadership practices.

A systematic review (SLR) of the research literature about executive leadership in the *child welfare sector in Australia* was conducted in order to better understand what constitutes effective leadership, with a focus on the executive level in Human Services Organisations (HSOs) and the leadership required to implement Evidence Based Programs (EBPs), which are shown to improve outcomes for clients of HSOs (McCarthy & Griffiths, 2021). This is a niche area, with very few studies specific to it (Bish & Becker, 2016; Young et al., 1993; McMurray et al., 2010; Sarros et al., 2011). For this reason, the research was extended to look at leadership and leadership implementation of EBPs in HSOs more broadly around the world.

While there has been much written about leadership, there is little consensus on a common definition of leadership, excepting many authors agree that leaders exert influence and provide direction (Covey, 2004; Simons, 1998). Influence needs to be referent and legitimate (Lunenburg, 2011, p.5) but cannot be useful without the leader taking action and shaping behaviour while leading others, often towards a common goal (Bass & Stogdill 1990). Leaders influence, motivate and contribute to the overall effectiveness of the organisations they lead (House et al. 2004, cited in Lamond 2009, p.359). Research on leadership commonly uses leadership and management interchangeably, with leadership often conceptualised as leadership (or management) support (Reichenpfader et al., 2015)¹. There is a degree of overlap between leadership and management both in terms of the processes and the people involved, and that both capabilities are needed for organisational effectiveness (Yukl & Lepsinger, 2005). This is true of the HSO and Non-Profit (NFP) organisational setting, which has historically demonstrated a more complex set of expectations regarding the importance of leadership versus managerial skills (Drucker 1990 and Hesselbein 2004, cited in (Lansford et al., 2010) than needed for for-profit organisations. Within HSOs leadership is important, from the official leadership of the Chief Executive Officer (CEO) right down to the informal leadership of citizens. Nevertheless, leadership is identified as requiring skills different from management (Kibort, 2004). Leaders are differentiated by their personal attitude toward organisational goals, approaching work as an opportunity for risk and reward, forming subjective relationships with 'followers',³ and having a sense of self-independence to the organisation employing them (Zaleznik, 1977).

3 Evidence-Based Decision-Making

Evidence-based decision making (EBDM) involves using multiple sources of data from reliable sources (for example, peer-reviewed literature) to inform decision making. It may require searching research databases, determining user statistics and/or surveying stakeholders and practitioners. Evidence-based decision making provides a solid basis and justification for management decisions.

However, a recent systematic review of over 5000 papers found that taking an evidence-based approach to decision making is difficult as a solitary pursuit (HakemZadeh & Rousseau, 2024). To be effective, according to this study, EBDM needs a community: people to communicate with. The three groups of stakeholders were Identified as "(a) researchers (internal or external

³ In the Australian context, 'followers' are more commonly referred to as 'workers'.

to the organization), (b) stakeholders (people in an organization or community who are directly or indirectly affected by a decision), or (c) communities of practice (people who have job responsibilities similar to those of the decision-maker". Without institutional support and resources, including an organisational climate emphasising science over politics, and a collaborative and collective culture, it is difficult to promote EBDM (HakemZadeh & Rousseau, 2024). This is why an organisation such as OzChild, which is promoting the use of EBDM and providing professional development training in the form of the Evidence-Based Decision Making for Human Services Leaders programme is so important to research, to determine how the understanding the principles of EBDM translates into leadership practice, in this case in a child welfare context.

These principles are broken down into six phases of EBDM and four pillars. The four pillars are:

“Evidence from the scientific literature: Findings from published empirical studies

Evidence from the organization: Data, facts, and figures gathered from the organization

Evidence from practitioners: The professional experience and judgment of practitioners

Evidence from stakeholders: The values and concerns of people who may be affected by the decision” (Barends et al., 2025 p. 11) .

The six phases of EBDM will be discussed in section 3.1 Theoretical Lens.

3.1 Theoretical Lens

The theoretical lens that underpins this research is evidence-based decision-making. This practitioner-led philosophy provides a framework which allows individuals to understand and critically evaluate scientific information to inform decision-making (Dawson et al., 2024).

The principles underlying EBDM are embodied in what is known as the four pillars:

Evidence from Scientific Literature

Evidence from the organisation

Evidence from stakeholders

Evidence from practitioners (Barends et al., 2025)

EBDM has six main phases: ask, acquire, appraise, aggregate, apply and assess (see Figure 1). These are known as the 6A's.

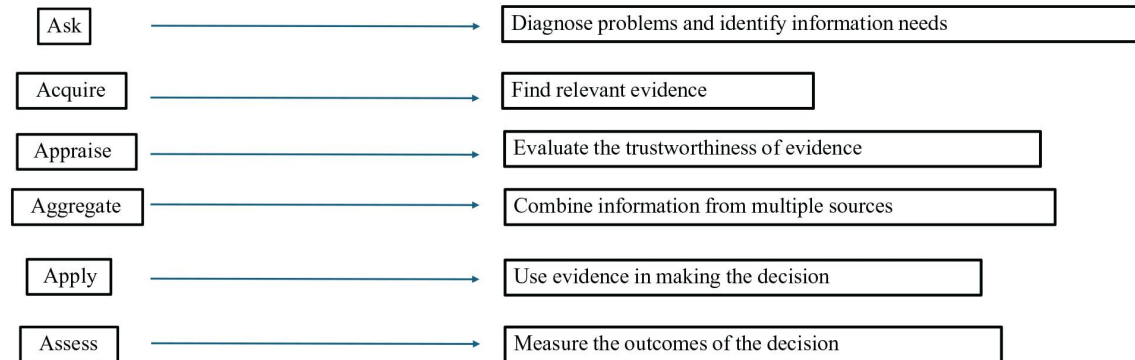


Figure 1. The six phases of EDBM (HakemZadeh & Rousseau, 2024)

The six phases of EDBM - the process that practitioners need to undertake to ensure that decision making is evidence-based, sits on four pillars of evidence. These four pillars are: organisational data, practitioners, stakeholder's voices and concerns and scientific literature (studies). Omitting any of these pillars of evidence means that a major element of evidence-based decision making is omitted. All four pillars, when undertaken within the process incorporating the six phases leads to managerial and leadership decisions that are science-based and are formulated to be a good fit in the context of individual organisations (Open Learning Initiative, 2025).

However, there are obstacles to evidence-based practice. In a 2017 study, 2789 management practitioners were surveyed on their attitudes towards evidence-based practice. A lack of time to review scientific literature, and a limited understanding of scientific research were cited as major barriers to implementing an evidence-based practice approach. Overall, managers were positive towards the idea of evidence-based practice (69%) but lacked detailed knowledge of the process required. Instead, 91% of managers based their decision making on prior experience rather than taking an evidence-based approach. Over 70% of managers were familiar with the term "Evidence-Based Practice" yet most failed to describe or gave limited answers to what evidence-based practice meant (Barends et al., 2017). This reinforces the need for institutional leadership training, cultural change, collaboration and resourcing. Taking an evidence-based approach requires a holistic approach and commitment on the part of organisations: It cannot exist in a vacuum.

The importance of this research is to measure to what extent leadership decision-making is changed when Evidence-Based Decision-Making (EBDM) principles are applied in professional practice, in this instance in the child welfare sector. Generally, leadership programs are reviewed employing once-only self-reported surveys, questionnaires or focus groups. These may be undertaken by the leaders themselves, or by fellow employees or followers who rate their leaders. The dependence on subjective verbal and textual reports raises a measurement and validity issue. Analysing verbal (and textual) data can lead to exhaustive examination of the relationships between words, which in itself is interesting, but does not necessarily investigate the nexus between words and behaviour: what people say is not necessarily what they do (Podsakoff & Organ, 1986). This leads to external validity issues, data may be difficult to validate, or not at all. Generalisability and predictability of the data therefore become

ineffective: verbal (and textual) data may not lead to accuracy in predicting actual behaviour. (Podsakoff et al., 2003).

The difficulty in this research was to develop a research methodology that afforded the analysis of thinking in both the participants and the Control group as well as validating this qualitative approach by integrating quantitative methods such as pre- and post-testing, employing a quasi-experimental approach (Capili & Anastasi, 2024; Zhu et al., 2020). This mixed methods approach allows analysis of the content in the form of codes via protocol analysis, examination of the attitudes of the participants via textual analysis to identify themes that arise in the data, as well as quantitative comparisons between the two cohorts and the broader EBDM student community.

There is a significant gap in the research using controlled comparison trials that measure leadership decision making before and after the intervention of a program. There is a need to examine the effectiveness of the Evidence-Based Decision Making for Human Services Leaders professional development programme to test the hypothesis that it improves decision-making in those leaders that undertake it. The qualitative findings are reinforced by determining if there is an improvement in the quantitative pre- and post-test scores by the participants. These scores are compared with a larger EBDM course cohort who are not executive leaders in the child welfare sector to determine if the test scores are indicative of the broader professional development course population.

This research also seeks to determine if the Control group's scores and "think aloud" sessions show signs of increased use of evidence-based decision-making, which would indicate cultural change within the OzChild organisation.

4. Research Approach and its Rationale

This research takes a constructivist approach as a research paradigm, specifically focussing on social constructivism. In this worldview, the participant is considered the main protagonist, that is, the central character in the obtaining of knowledge. It states that knowledge is socially constructed. Learning can be individually cognized; however, learning is subsequently validated through social interaction (Doolittle & Hicks, 2003). So, learning can come from a classroom, for example, but this learning is validated by comparing the knowledge obtained with real-world observations and in practice. Learning is therefore validated by seeing how that knowledge is used and applied in broader society, particularly in practice. These two arms of constructivism, individual and social are taken to be interconnected, and both contribute meaningfully to learning and knowledge acquisition (Álvarez-Correa & Quintero-Arrubla, n.d.).

One of the areas of interest in this research is to ascertain if by increasing the awareness of an evidence-based approach in an organisation, whether individuals who have not undertaken training increase their use and understanding of using evidence for decision making (see Research Objective 1). Informal learning is considered part of social constructivism. It is defined as learning that does not take place in a formal setting, but occurs spontaneously

through collaboration, interaction and imitation. This aligns closely with the principles of social constructivism.

Social constructivism takes a pragmatic approach, it involves learning by doing. Pragmatism is also a guiding paradigm in this research study. Pragmatism is a problem-centred approach closely associated with mixed-methods research. It bridges "the gap between the scientific method and structuralist orientation of older approaches and the naturalistic method and freewheeling orientation of newer approaches" (Creswell, 2013; Kaushik & Walsh, 2019).

The choice of a quasi-experimental design is driven by practical and ethical considerations. Random assignment of participants to different groups is not feasible or ethical in this context. By using a quasi-experimental research design, we will compare the outcomes between the intervention group from OzChild (those who receive the Evidence-Based Decision Making for Human Services Leaders program) and the Control group (those who are leaders in the OzChild organisation) who have not undertaken the professional development. This design allows for a comparison of the two groups while accounting for some potential confounding variables. The final group for evaluation consists of a larger cohort of participants in the Evidence-Based Decision Making for Human Services Leaders program who are not from the OzChild organisation.

4.1 Research Design

This research employed a mixed methods Quasi-Experimental research design and collected quantitative and qualitative data across three research stages:

- Stage one involved a pre-intervention "think aloud" session for both groups using a unique scenario which presented situations likely to occur in professional practice.
- Stage two involved pre- and post-tests prior to the course cohort undertaking their training and afterwards. Both the intervention group and the control undertook both sets of tests.
- Stage three involved a post-course "think aloud" session using a different scenario, also based on professional practice, again undertaken by both the intervention group and the control.

The "think aloud" sessions were analysed using qualitative protocol analysis, with cross-coder checking using Krippendorff's alpha. Insightful quotes and trends are also reported. The quantitative pre- and post-test results were then compared, and with a larger cohort who had undertaken the EBDM professional development course, who were not necessarily members of senior executive leadership teams.

This research employed qualitative "think aloud" protocol analysis and grounded research methods, identifying codes/themes and quotes in conjunction with quantitative research in the form of pre- and post-intervention tests. As the final number of participants who completed the study (n=10) was small, their quantitative pre- and post-test results are compared with a larger cohort of participants who completed the EBDM course. This study is a longitudinal study, with the EBDM professional development course for leaders in the child

welfare sector in Australia taking on average 8 months, with two intakes six months apart. The research was conducted in 2025.

In addition to the tests and "think aloud" protocol analysis and grounded research, a demographic survey was also undertaken to determine the age, gender and professional backgrounds of the participants, as prior knowledge and training could influence the results.

Figure 2 shows a diagrammatical representation of the research design. The mixed-methods research methodology required careful choice of participants: not all employees of OzChild qualified for either the course or the study. The participants were divided into two groups: those who undertook the Evidence-Based Decision Making for Human Services Leaders program and those who did not. It is expected that the Control group should show differences, given that they did not undertake the professional development course.

The larger cohort of participants in the course who were not members of the OzChild organisation were used for comparison with the participants' results to illustrate the differences in the quantitative analysis of the pre- and post-test results. These participants will be compared with the standard deviations of the cohort to determine if there are differences between them and the OzChild professional development cohort, given the small sample.

An online survey was completed by all participants from OzChild to gather demographic information. This included gender, age, years of work experience, qualifications and experience in areas such as leadership, human resources, retail, sales and marketing, psychology, among others. This information was gathered to determine the background to the participants to set the context for the baseline results.

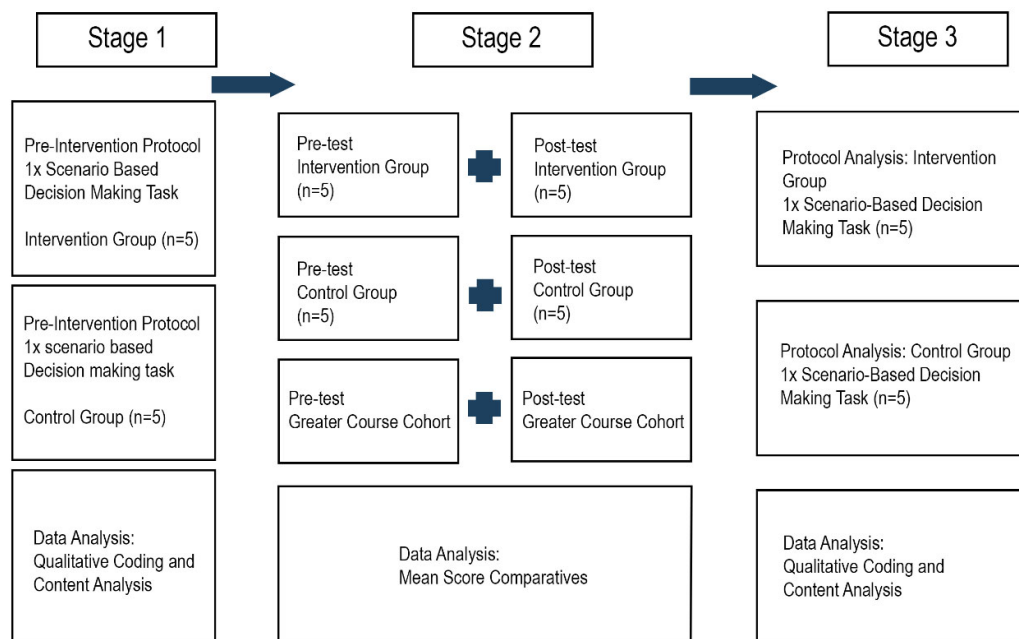


Figure 2 Quasi-Experimental mixed methods research design

Stage 2

Pre- and Post-Test Assessments: The groups underwent pre- and post-test assessments. The rationale behind this choice was to measure the intervention participants' decision-making abilities both before and after the training program, in a professional context. Pre-test assessments establish a baseline of participants' decision-making skills, while post-test assessments capture any changes or improvements resulting from the training interventions. The inclusion of the Control group in the pre and post-tests provides a comparison with the EBDM course group.

Stages 1 & 3

Qualitative Technique-Protocol Analysis: To evaluate the decision-making abilities of the participants, the qualitative technique of protocol analysis has been employed, using the scenario-based decision-making tasks. This method involves presenting decision-making scenarios or tasks to the participants and collecting their "think aloud" responses, observations, and reasoning processes in real-time. The rationale behind using protocol analysis is to gain insights into participants' decision-making strategies, thought processes, and problem-solving approaches. Each "think aloud" session was timed for a maximum of 15 minutes, although many participants finished prior to this duration. The "think aloud" sessions were then coded using codes (or themes) to determine the frequency each code was used. Cross coder reliability was employed using Krippendorff's Alpha (Krippendorff, 2011). The frequency of each code is then summed to determine if there is an increase in occurrence between the pre-intervention "think aloud" scenario session and the post-intervention "think aloud" session. This would sufficiently answer research questions 1, 1a, 1b and 2. Qualitative quotes are used to illustrate each code (theme). Reporting of the results integrates both the qualitative and quantitative data to gain insights into the data.

Scenario-Based Decision-Making Tasks: Participants from both the intervention and control groups have been presented with scenario-based decision-making tasks during the protocol analysis. These scenarios have been designed to reflect real-world challenges and dilemmas commonly encountered in management decision-making. The rationale for this approach is to assess participants' decision-making skills in situations relevant to their work context and to capture their application of learned principles and strategies from the training programs. Due to government protocols and regulations surrounding management decision-making in the child welfare context, it was determined to use more general leadership decision-making scenarios rather than those specifically concerning children or child welfare.

Data Analysis: The data collected from pre- and post-test assessments and the protocol analysis stages was analysed using appropriate statistical and qualitative analysis techniques. These results have been analysed and compared with results from a larger cohort of participants of the Evidence-Based Decision Making for Human Services Leaders program.

Stage 2: Quantitative analysis allows for the comparison of mean scores or changes in decision-making outcomes between the pre and post-tests for both OzChild cohorts and the

larger course cohort, who are not from OzChild. Results will be reported according to qualification level and leadership level.

Stage 1 & 3: Qualitative analysis of the protocol data involved coding, categorizing, and identifying patterns and themes in participants' decision-making processes and reasoning.

4.2 Ethics Clearance

This research was approved by the Torrens University Australia Ethics Board.

Ethics Application 0381 Examining the Impact of Evidence-Based Decision-Making Professional Development in the Child Welfare Sector (see Appendix 11.1)

5. Analysis and Outcomes

Participants

The participants from the OzChild organisation were carefully chosen for participation in this research. All participants were situated in Australia. Firstly, the participants had to come from senior leadership of the organisation. Of those invited to participate, half were enrolled in the Evidence-Based Decision Making for Human Services Leaders program. None of the participants had been involved in Evidence-based decision-making professional development previously. The participants in the professional development program were enrolled for 12 months, but most completed the online course after eight months of participation. Participants from two intakes were included in the study: from May 2024 to May 2025 and November 2024 to November 2025.

This longitudinal quasi-experimental mixed methods study occurred between the dates mentioned above. Initially 14 participants were recruited for the study. Of those, one resigned from the organisation, two did not continue with the course and to keep the Control group and intervention group equal numbers, one was not invited to participate. This left ten participants in all, five from each group.

Of the ten participants, the majority were women. The Control group comprised three women and two men. The intervention group comprised four women and one man.

The ages of the participant were towards the older age range (four participants each for the following two age ranges (40-49 and 50-59 years of age), which is expected of a participant group comprising senior leadership. There was one participant each in the following two age ranges (20-29 and 30-39 years of age). There were no participants over 60 years of age

Similarly, the years of work experience also reflected the older age range of the participants (see Table 1). Interestingly, the participant with the least work experience had completed a master's degree.

Table 1 Years of work experience

Years of Work Experience	
0-2	
3-5	1
6-10	1
11-20	3
Over 20	5

The most frequently acquired post-secondary qualifications were a bachelor's degree, Certificate IV and Certificates 1-3. Some participants had multiple qualifications (see Table 23).

Table 2 Participant's post-secondary qualifications

Post-Secondary Qualifications (choose all that apply)	
Doctoral degree	
Master's degree	2
Graduate diploma	1
Bachelor's Degree	3
Associate degree	
Diploma	
Certificate IV	3
Certificate 1-3	3
Other	Advanced Diploma (2)

Human resources experience was the most frequently occurring work experience, followed by leadership of business units (see Table 3).

Table 3 Previous work experience of participants

Previous Experience (choose all that apply)	
Human Resources	5
Human Resource Development	3
Leadership of Business Units	4
Retail	3
Sales	2
Marketing	
Product Innovation	1

Psychology	2
Psych Nursing	
Allied Health	2
Other	Social work, Social Services, Information Technology

Most of the participants were in Melbourne (n=6), with the second most frequent place of work being Brisbane (n3) and one participant from Sydney.

Inter Coder Reliability

One issue that arose with the "think aloud" data was classifying when the participants were investigating whether the premise as presented in the scenario was valid or not. This often involved re-reading the scenario and questioning the characters' use of evidence-based thinking. One coder put this in the "Asking" category, and another in the "Appraising" category. It was decided that only questions would go into the "Asking" category, otherwise when the participant appraised the overall premise of the scenario, and it would go into the "Appraising" category. For this reason all the "think aloud" transcripts were re-coded to reflect this change. This change was enacted prior to doing the cross-coder reliability check.

The protocol analysis coding was validated using the cross-coder reliability check Krippendorff's alpha. Krippendorff's alpha (α) is a reliability coefficient between different coders. It is used to question how much the data generated can be trusted, given that the evaluation between the coders is subjective. It can be used for different types of data: nominal, ordinal, ratio and interval data (Krippendorff, 2011). In this case the data will be nominal, representing a yes/no agreement between the coders. Because there are so many categories/codes, this is considered a multi-label analysis. The nominal data has 12 codes asking, acquiring, search, organizational data, appraising, scientific data, aggregating, stakeholders, applying, culture, assessing and practitioners. These codes emerged from the professional development course content and a set of keywords derived from the "think aloud" data (see Appendix 11.4 Keyword Lists).

In this check a resulting α value of 1 denotes perfect agreement whereas a result of 0 indicates an absence of reliability. Its general form is:

$$\alpha = \frac{p_o - p_e}{1 - p_e}$$

Figure 3 Krippendorff's basic formula (Kaplan, 2025)

Where p_o is the observed disagreement between coders and p_e is that which would be attributable to chance, otherwise known as the expected agreement (Kaplan, 2025; Krippendorff, 2011). A value of $\alpha \geq 80$ is considered a satisfactory level for alpha, a value of 67-79 is the lower bound of acceptable values, and less than that is considered poor (Marzi et al., 2024).

In this case, four "think aloud" transcripts were subject to Krippendorff's alpha: the pre-intervention "think aloud" data - one from the Control group and one from the intervention group. The post-intervention data - one was used from each group, chosen randomly. Out of a total of 20 "think aloud" transcripts this was deemed sufficient to indicate reliability of the coding. Two coders were used.

Calculating multi-label binary Krippendorff's alpha is considered complex, for this reason computer-based methods are employed generally. Several tools exist, including R, K-alpha and NVIVO. Unfortunately for multi-label binary data with two coders, the format required for K-alpha does not include multi-label data. Another option is to use ChatGPT, which is free and open access. For this analysis only, ChatGPT was employed to determine the Krippendorff alpha value.

The nominal multi-label data was cleaned to eliminate null data (where no category/code was chosen) and data with less than two choices (minimum of one per coder). Krippendorff's alpha can only be calculated with two or more values, one per coder, in this case.

The Krippendorff alpha calculations returned an average result of 70% agreement (see Appendix 11.6), which is considered satisfactory for multi-label data.

Protocol Analysis of the "think aloud" data

As previously stated, the "think aloud" data was categorized using 12 categories. The coding was guided by keywords that arose out of the course materials and the "think aloud" data (see Appendix 11.4 Keyword Lists). Because the participants and the course were tailored to those working in the child welfare sector where the word "practitioner" is used to describe practicing employees and one of the scenarios was not in a medical or child welfare context, it was decided to differentiate between managers (stakeholders) and practitioners (employees, therapists or team members).

Similarly, the coding of "Asking" and "Appraising" was clarified (see Inter coder reliability above). This deviates slightly from the Evidence-Based decision-making protocol in that "Appraising" in the EBDM context is meant to mean appraising the available data. What the participants spent a lot of time on was appraising the legitimacy of the scenario premise itself (i.e. questioning the fictional character's decision making and why they came to the stated conclusions). This meant that "Appraising" scores included both appraising the validity of the scenario as well as appraising the data, which accounts for the high values for this category.

Pre-intervention Results

Overall, the differences between the pre-think aloud of the Control group and the course cohort was not markedly different (see Figure 4). The Control group indicated slightly more in the "Asking" code and in the "Scientific Data" code. As this is multi-label coding, there are more codes than utterances, as there can be more than one code per utterance. The number of utterances is detailed in tables 4 and 5.

Table 4 Pre-Intervention Control group "think aloud" results, raw values only.

Pre-intervention Control group - raw values														
ID	Asking	Acquiring	Searching	Org Data	Appraising	Scientific Evidence	Aggregating	Stakeholders	Applying	Culture	Assessing	Practitioners	Total Codes	Utterances
P1	16	11	0	10	15	2	5	23	5	15	0	11	113	45
P3	7	6	2	2	21	4	0	8	1	6	0	7	64	43
P5	1	11	0	2	21	5	2	7	1	7	0	9	66	49
P7	19	8	0	7	18	0	0	9	2	3	2	12	80	51
P9	8	6	0	5	4	0	1	6	6	3	3	4	46	26
TOTAL	51	42	2	26	79	11	8	53	15	34	5	43	369	214

One participant, P7 overall had a high number of questions, concentrated at the beginning of the "think aloud" where of the first 10 utterances, six of them were questions. For example: "So the HR director feels that the company's workforce is aging and what he attributes rather than actually asking, I guess what, what's the evidence?"

Table 5 Pre-intervention Course cohort "think aloud" results, raw values only.

Pre-intervention Course group - raw values														
ID	Asking	Acquiring	Searching	Org Data	Appraising	Scientific Evidence	Aggregating	Stakeholders	Applying	Culture	Assessing	Practitioners	Total Categories	Utterances
P4	0	11	1	3	19	0	3	21	3	10	1	7	79	43
P8	8	6	0	6	22	1	6	11	0	8	0	11	79	47
P10	22	9	0	2	11	0	6	12	1	9	0	10	82	47
P12	5	11	1	8	4	0	4	4	1	3	0	6	47	25
P14	6	9	0	5	10	2	5	4	1	9	0	8	59	36
TOTAL	41	46	2	24	66	3	24	52	6	39	1	42	346	198

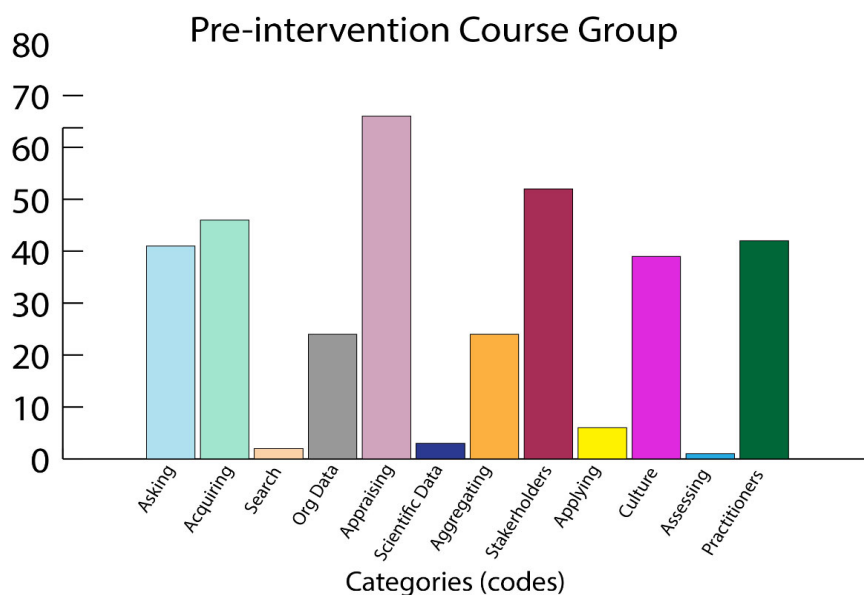
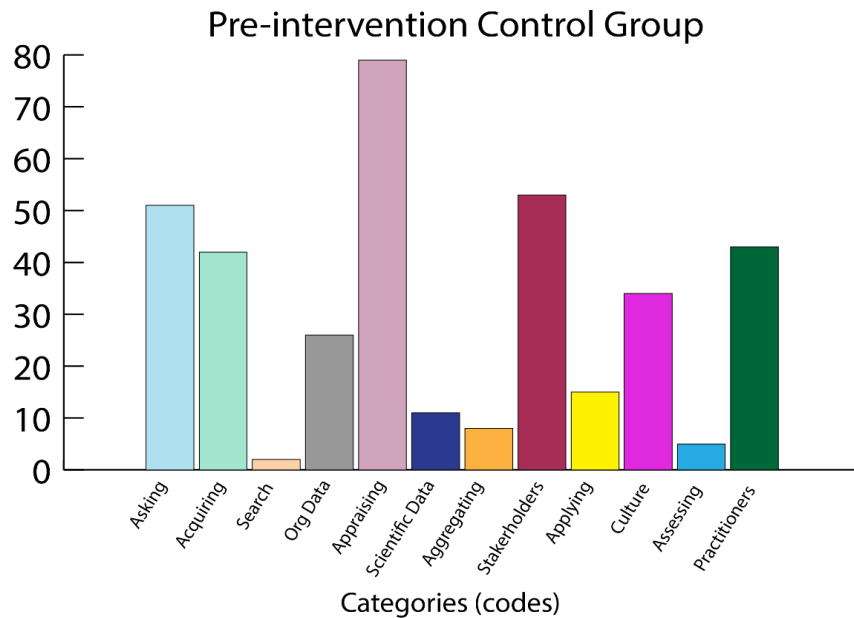


Figure 4 Pre-intervention "think aloud" bar charts, raw values only.

Another difference was in the "Aggregating" category, which was more evidence in the pre-intervention course group. This was evident in a comment from P8, who was focused on comparing the scenario's HR data with other companies and society.

" So, if we're looking at our particular company, if I'm working with Aziza, we need to look at other based organizations or potentially any research or articles that are based in science and can be proven."

The Control group also focused more on "Applying" a chosen result, be it based on their own investigations, or that proposed by the organisation's executives in the scenario. This was

indicated by P9, who, after investigating an evidence-based approach, still agreed with the approach proposed by the scenario's executives:

" Yeah, I think the solution would still be the same that you know, there is obviously evident there's a problem that they've identified, so it requires a solution, but that would be one solution I would think about, I think even at a potential survey of the team leaders within the organization."

Other than these differences, the overall trends in the results between the two groups appear consistent.

Given that each participant had a varying number of utterances, some with more than 50% more than others, displaying the results as percentages of total utterances indicates the relative frequency of the codes (see Figure 5). There appears to be very little difference between Figure 4 and Figure 5, except for the "Assessing" code normalizing to zero due to low numbers. Tables 6 and 7 provide the normalised results as percentages for the pre-intervention "think aloud" data.

Table 6 Pre-intervention Control group, percentages of participants' codes

Pre-intervention Control group - percentages													
ID	Ask ing	Acqui ring	Se arc h	Org Data	Appr aising	Scie ntific Evid ence	Aggr egating	Sta keh olders	Ap plying	Cul ture	Asses sing	Practiti oners	Tot al Co des
P1 %	14	10	0	9	13	2	4	20	4	13	0	10	113
P3 %	11	9	3	3	33	6	0	13	2	9	0	11	64
P5 %	2	17	0	3	32	4	0	11	2	11	0	14	66
P7 %	24	10	0	9	23	0	0	11	3	4	3	15	80
P9 %	18	13	0	11	9	0	8	13	11	7	7	9	46
	69	59	3	35	110	12	12	68	22	44	10	59	369
Weighted Average	14	11	1	7	22	2	2	14	4	9	2	12	100

Table 7 Pre-intervention EBDM course group, percentages of participants' codes.

Pre-intervention Course group - percentages													
ID	Ask ing	Acqui ring	Se arc h	Org Data	Appr aising	Scie ntific Evid ence	Aggr egating	Sta keh olders	Ap plying	Cul ture	Asses sing	Practiti oners	Tot al Co Des
P4 %	0	14	1	4	24	0	4	27	4	13	1	9	79
P8 %	10	8	0	8	28	1	8	14	0	10	0	14	79
P10 %	27	11	0	2	13	0	7	15	1	11	0	12	82
P12 %	11	23	2	17	9	0	9	9	2	6	0	13	47
P14 %	10	15	0	8	17	3	8	7	2	15	0	14	59
TOTAL	58	71	3	39	91	5	36	70	9	55	1	61	346

Weighted Average	12	14	1	8	18	1	7	14	2	11	0	12	100
------------------	----	----	---	---	----	---	---	----	---	----	---	----	-----

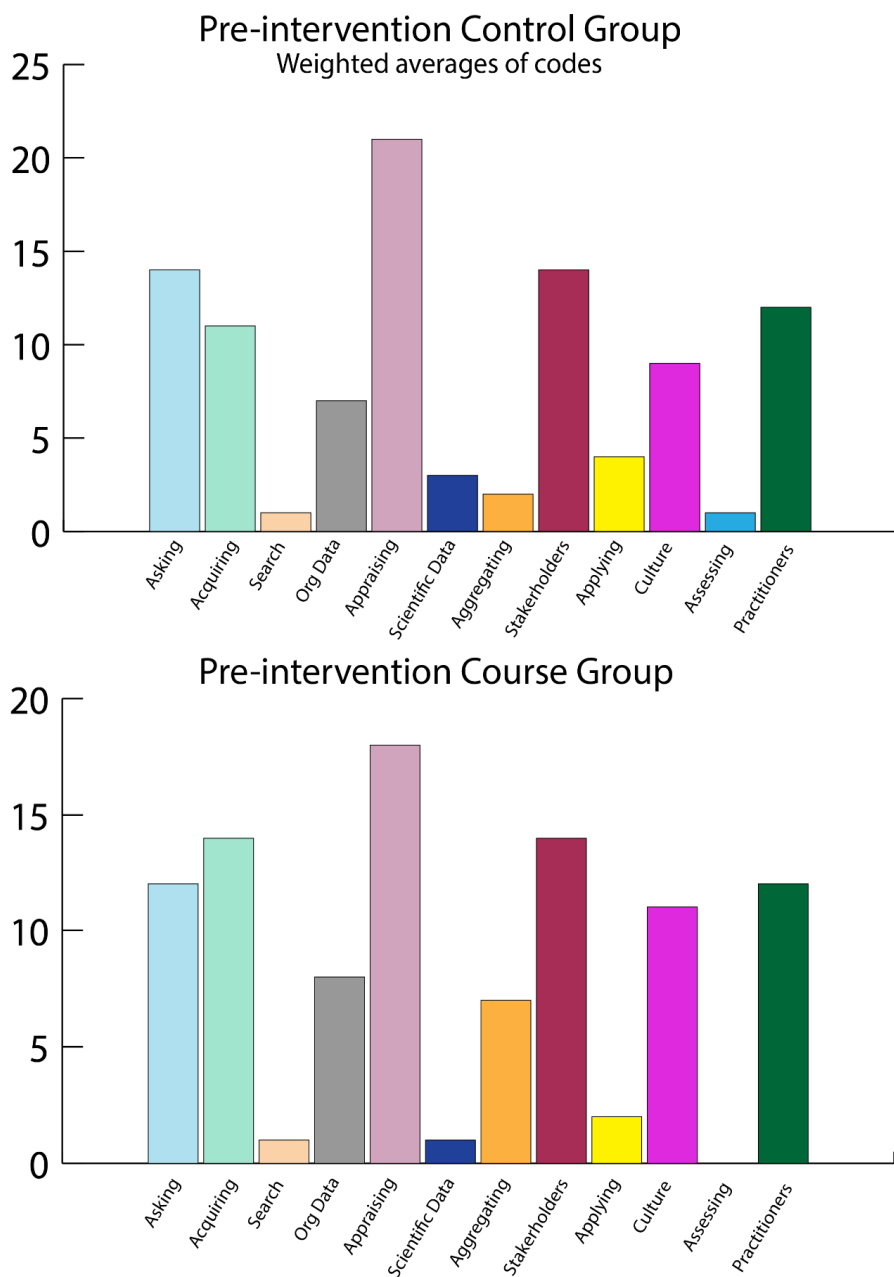


Figure 5 Weighted averages of participants' codes for the pre-intervention groups shown as percentages.

Post- Intervention Results

The participants who undertook the Evidence-Based Decision Making for Human Services Leaders professional development programme had eight months to complete the programme (or course). The programme was online and the participants could complete it in their own time, so there was no set completion date, but it did have to be completed at the eight-month mark. Two participants withdrew from the course, which left five active participants. Of the Control group, one resigned from the OzChild organisation and it was decided not to include a sixth to keep the numbers of participants and Control group the same.

There were two intakes included in this study, the second intake commenced in July 2025 and included one participant who qualified for this study. This participant completed the course within three months, so this person's "think aloud" occurred shortly after completion, which could result in higher overall scores.

The second "think aloud" exercise employed a different scenario from the first. What is interesting in these results is the complete absence of the use of scientific data in the Control group, which is interesting, as they did consider it more frequently in the pre-intervention "think aloud" scenario (see Figure 6). This is a marked difference between the two groups, where the Control group had zero examples of scientific data and the EBDM course group had 11 mentions, compared with 3 from the pre-intervention "think aloud" course group, compared with 11 mentions with the Control group. This is a reversal in the use of scientific data when examining the scenarios.

The Control group mentioned organisational data more frequently than the course cohort, but the course participants spent more time appraising the evidence and scenario. The Control group had 31 "Appraising" utterances, compared with 42 in the EBDM course cohort. Once course participant P14 was quite high in the "Appraising" category, both of the solution outlined in the scenario and of the evidence:

"I would be obviously then appraising the information that I have. I would have to look at all of the intersecting, well biases I think is the word, and what they're wanting out of it, but also managing expectation."

Table 8 The post-intervention control group "think aloud" results - raw numbers

Post-intervention Control Group - raw values														
ID	Ask ing	Acq uiri ng	Se arc h	Org Dat a	Ap prai sin g	Scie ntific Evid ence	Ag gre gati ng	Stak ehold ers	Appl ying	Cul ture	Asse ssing	Pra ctiti one rs	Total Cate gories	Utter ances
P1	8	4	0	20	9	0	0	7	0	0	0	5	53	28
P3	5	8	2	14	7	0	6	6	4	0	3	2	57	32
P5	7	10	1	8	13	0	3	6	0	1	0	9	58	32
P7	19	10	0	9	2	0	0	4	0	1	1	8	54	29
P9	11	13	0	14	0	0	9	17	3	1	0	8	76	49
TOT AL	50	45	3	65	31	0	18	40	7	3	4	32	298	170

Table 9 The post-intervention course group "think aloud" results in raw numbers

Post-intervention Course Group - raw values														
ID	Asking	Acquiring	Searching	Organising	Appraising	Scientific Evidence	Aggregating	Stakeholders	Applying	Culture	Assessing	Practitioners	Total Categories	Utterances
P4	4	8	1	3	8	1	2	11	0	0	0	6	44	29
P8	15	5	1	3	9	3	2	8	0	0	0	7	53	38
P10	10	7	2	7	7	1	5	8	1	1	1	7	57	34
P12	1	2	0	4	5	2	1	4	0	0	0	5	24	14
P14	5	7	1	5	13	4	8	8	2	0	1	5	59	34
TOTAL	35	29	5	22	42	11	18	39	3	1	2	30	237	149

The total number of utterances in the post-intervention "think aloud" activities was less for both groups than in the previous "think aloud" session, as outlined in tables 8 and 9, in comparison with tables 6 and 7. These tables correspond with Figure 6 (below).

To gain an overview of the relative frequency of the results, again they need to be normalised as percentages, this allows for statistical evaluation of the results. It also allows more exact comparison of the frequency of the utterings overall. See Table 10 and

Table 11 for results shown in percentages.

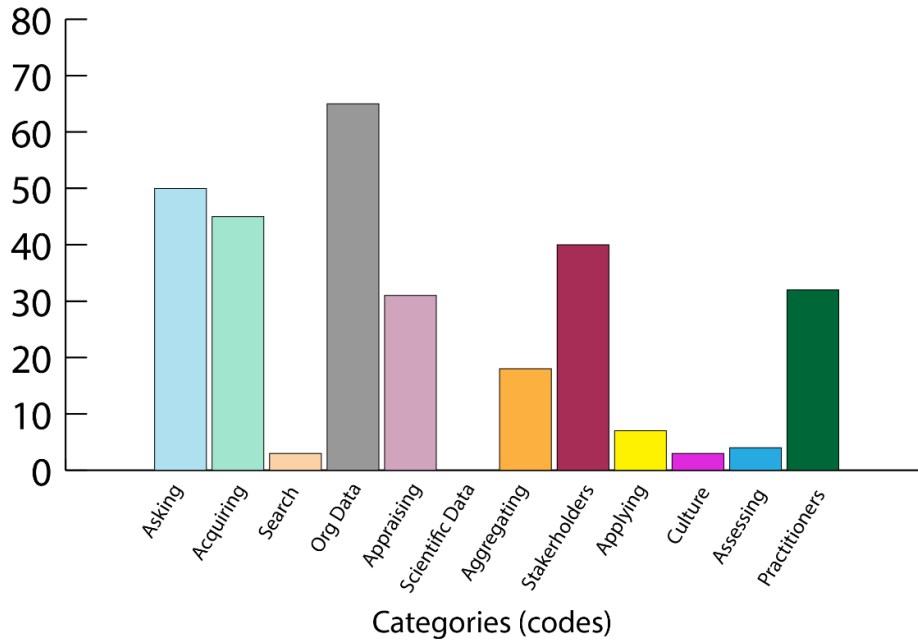
Table 10 Post-intervention Control group shown in percentages

Post-intervention Control Group - percentages													
ID	Asking	Acquiring	Searching	Organising	Appraising	Scientific Evidence	Aggregating	Stakeholders	Applying	Culture	Assessing	Practitioners	Total Codes
P1 %	15	8	0	38	17	0	0	13	0	0	0	9	53
P3 %	9	14	4	25	12	0	11	11	7	0	5	4	57
P5 %	12	17	2	14	22	0	5	10	0	2	0	16	58
P7 %	35	19	0	17	4	0	0	7	0	2	2	15	54
P9 %	14	17	0	18	0	0	12	22	4	1	0	11	76
TOTAL	86	74	5	111	55	0	28	64	11	5	7	54	298
Weighted Av %	17	15	1	22	11	0	6	13	2	1	1	11	

Table 11 Post-intervention course group shown in percentages.

Post-intervention Course Group - percentages													
ID	Ask ing	Acqu iring	Se arc h	Org Data	Appr aisin g	Scientific Evidence	Aggre gating	Stake holde rs	Ap plyi ng	Cul tur e	Asse ssin g	Pract ition ers	Total Codes
P4 %	9	18	2	7	18	2	5	25	0	0	0	14	44
P8 %	28	9	2	6	17	6	4	15	0	0	0	13	53
P10 %	18	12	4	12	12	2	9	14	2	2	2	12	57
P12%	4	8	0	17	21	8	4	17	0	0	0	21	24
P14%	8	12	2	8	22	7	14	14	3	0	2	8	59
TOTAL	68	60	9	50	90	25	35	84	5	2	3	68	237
Weighted Av %	14	12	2	10	18	5	7	17	1	0	1	14	

Post-intervention Control Group



Post-intervention Course Group

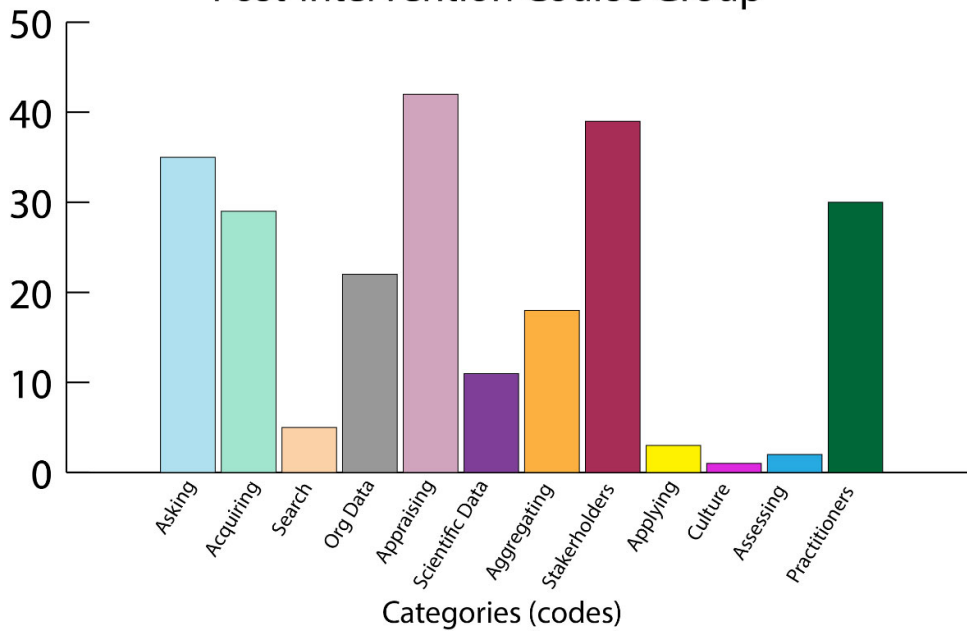
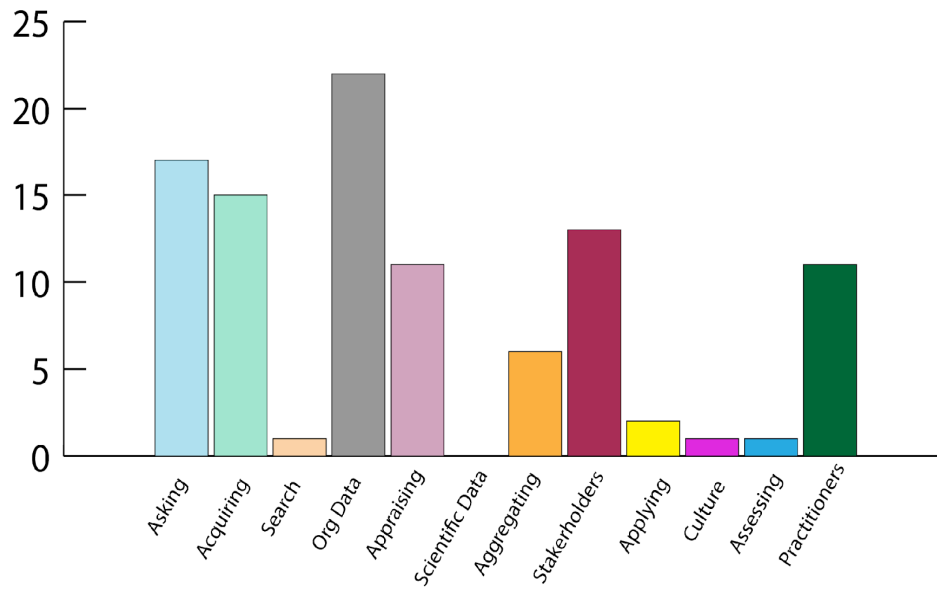


Figure 6 Post-intervention results *note scale on left hand side is different*. Note these are the raw figures.

Figure 6 shows the difference in the two groups after the intervention. As mentioned previously the intervention group shows more appraising than use of organisational data, the opposite is true of the Control group. Consideration of workplace culture disappears entirely for the intervention group. In the intervention group there is more consideration of "Stakeholders" and "Practitioners" than in the Control group, as well as more consideration of "Scientific Data". These all align with the tenets of Evidence-based decision-making. The

participants who completed the EBDM professional development program also showed more use of "Search" but less of organisational data, as an overall proportion of their thoughts. Figure 7 shows the codes normalised as percentages of utterances.

Post-intervention Control Group Weighted averages of codes



Post-intervention Course Group

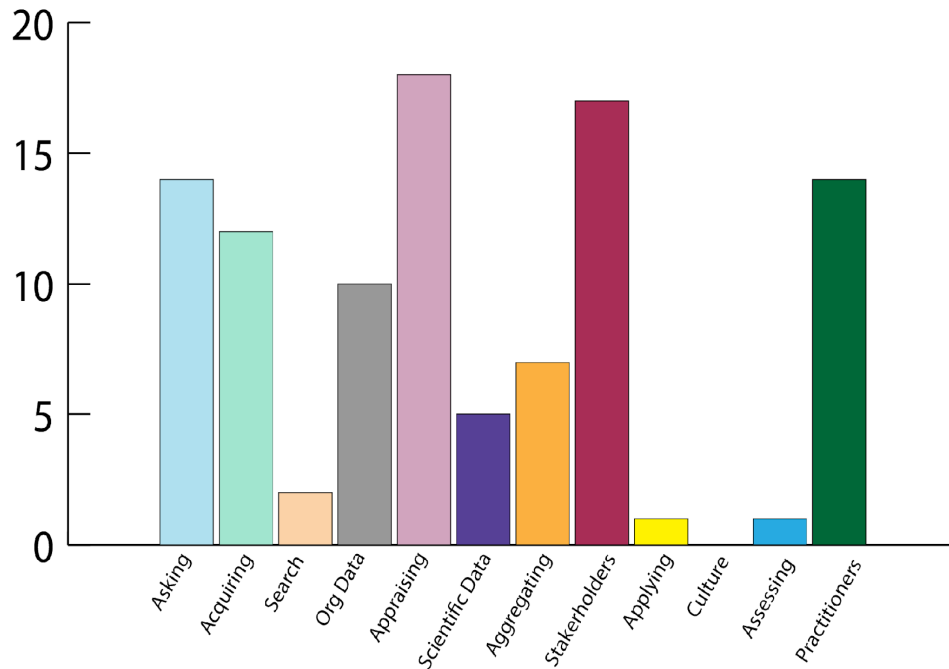


Figure 7 Graphs of normalised "think aloud" results shown as percentages.

Statistical Analysis of the Cohort Data

It is possible to do a paired two tailed T test to determine if there is statistical significance between the data sets. The paired two tailed T test is used to measure the same subjects twice, once before the intervention and again afterwards. It is preferable to use raw values for a paired two tailed t test, to reveal nuances in the sample size and variance in the values.

To answer **RQ1**: To what extent does the Certificate in Evidence-Based Decision-Making for Human Service Leaders produce changes in the application of evidence-based principles and practices in decision-making among executive leaders in the child welfare sector?

Course Participant Group T-Test Results

This test determines if the differences in participants who completed the EBDM for leaders in the child welfare sector professional development course changed from the pre-"think aloud" exercise and the post-"think aloud" exercise. A p value of 0.05 is sufficient to reject the null hypothesis. For the p value, a score nearer to zero is considered more significant. A t value above 2 is considered statistically significant at the $p \leq 0.05$ level (Social Science Statistics, 2026).

When using the raw values when comparing the pre "think aloud" results from the participants who completed the Certificate in Evidence-Based Decision-Making for Human Service Leaders before and after their course, a T-test calculation for 2 Dependent means returned a statistically significant result.

Table 12 Pre and post frequencies as raw values for course group.

Pre	41	46	2	24	66	3	24	52	6	39	1	42
Post	35	29	5	22	42	11	18	39	3	1	2	30

The values returned a p value of 0.031. The Cohen's d value was 0.71, which is considered a medium effect. The T value was 2.4660 (Social Science Statistics, 2026). This is considered a statistically significant result. However, what these values indicate is a statistically significant *reduction* in the number of responses. There were significantly more responses in the first "think aloud" session than the second.

Table 13 Paired results from the course group raw results show a significant decrease in the number of utterings. Pre "think aloud is X_1 and post X_2 .

Pair	X_1 (Pre)	X_2 (Post)	$D = X_1 - X_2$
1 (Asking)	41.00	35.00	6.00
2 (Acquiring)	46.00	29.00	17.00
3 (Search)	2.00	5.00	-3.00

Pair	X ₁ (Pre)	X ₂ (Post)	D = X ₁ - X ₂
4 (Org data)	24.00	22.00	2.00
5 (Appraising)	66.00	42.00	24.00
6 (Scientific data)	3.00	11.00	-8.00
7 (Aggregating)	24.00	18.00	6.00
8 (Stakeholders)	52.00	39.00	13.00
9 (Applying)	6.00	3.00	3.00
10 (Culture)	39.00	1.00	38.00
11 (Assessing)	1.00	2.00	-1.00
12 (Practitioners)	42.00	30.00	12.00

Employing the raw values from the five course participants we can see the breakdown of results.

Table 14 Breakdown of individual two-tailed paired T-test results course group (Social Science Statistics, 2026)

Participant	P value	Cohen's d	T value	Significance
P4	0.061	0.60 (medium effect)	2.0906	Not significant
P8	0.054	0.62 (medium effect)	2.1568	Not significant
P10	0.145	0.45 (small effect)	1.5684	Not significant
P12	0.045	0.65 (medium effect)	2.2603	Significant
P14	1.00	0.00 (negligible)	0.000	Not significant

What this data indicates is that the number of utterings from the course group was significantly less in the second “think aloud” session over the first, but on examination of Table 13 the number of utterings in the “Scientific data” code has increased, but not significantly. The greatest reduction in utterings was from P12, who ‘drew a blank’ on during the second “think aloud” activity. This influenced the overall score.

Control Group T-Test Results

The Control group did not undertake the Certificate in Evidence-Based Decision-Making for Human Service Leaders and had previously not undertaken training in evidence-based decision-making.

Using the raw results as a group:

Table 15 Pre and post frequencies as raw values for the Control group.

Pre	51	42	2	26	79	11	8	53	15	34	5	43
Post	50	45	3	65	31	0	18	40	7	3	4	32

This raw data returned a non-significant result. The p value was 0.355, the Cohen's d was 0.28 (a small effect) and the T value was 0.9659. This result did not satisfy the null hypothesis and is not statistically significant.

Table 16 Paired differences in the Control group pre (X_1) and post (X_2) "think aloud" data

Pair	X_1 (Pre)	X_2 (Post)	$D = X_1 - X_2$
1 (Asking)	51.00	50.00	1.00
2 (Acquiring)	42.00	45.00	-3.00
3 (Search)	2.00	3.00	-1.00
4 (Org data)	26.00	65.00	-39.00
5 (Appraising)	79.00	31.00	48.00
6 (Scientific data)	11.00	0.00	11.00
7 (Aggregating)	8.00	18.00	-10.00
8 (Stakeholders)	53.00	40.00	13.00
9 (Applying)	15.00	7.00	8.00
10 (Culture)	34.00	3.00	31.00
11 (Assessing)	5.00	4.00	1.00
12 (Practitioners)	43.00	32.00	11.00

There was therefore no measurable difference in the number of utterings between the pre and post intervention in the "think aloud" data for the group as a whole. However, the data shows a significant increase in the use of the use of Organisational data in Table 16.

Table 17 Breakdown of individual two-tailed paired T-test results Control group (Social Science Statistics, 2026)

Participant	P value	Cohen's d	T value	Significance
P1	0.014	0.73 (medium effect)	2.5192	Significant
P3	0.376	0.09 (negligible effect)	0.3236	Not significant
P5	0.298	0.16 (negligible effect)	0.5462	Not significant
P7	0.075	0.45 (small effect)	1.5508	Not significant
P9	0.028	0.62 (medium effect)	2.1458	Significant

The individual Control group's responses show that two participants did achieve statistically significant results between the two "think aloud" sessions. Both Participant 1 (P1) and Participant 9 (P9) showed statistically significant results between the pre- and post-intervention results, also indicating a reduction in utterings. This shows that there is a medium effect on some members of the Control group when their colleagues undertake EBDM training. However, this is more than offset by the negligible to small results in the other Control group members. Examining the utterings in more detail should be able to determine in what areas these changes occur, which will be examined in the next section.

Statistical Analysis of Larger Course Population

The statistical analysis of the larger EBDM course population included participants who were not part of the OzChild organisation, nor necessarily from senior leadership positions. These participants undertook pre-course tests and in-course tests. They did not participate in the later post-course tests, which in some instances were months after the completion of the course, therefore there are no results from this larger cohort that indicate the effect of the Forgetting Curve (Murre & Dros, 2015). These results therefore indicate the increase in scores that were gathered before and during the course, or immediately afterwards. The total population of OzChild employees (n=27) was used for this analysis, not just those involved in this more detailed study.

Population

The participants in this analysis are bachelor and master students from universities and business schools in the USA, Canada, Australia, Belgium, the Netherlands, and Germany. In addition, executive students (MBA, MSc) and professionals enrolled in organisational professional development programs in the USA, Australia, Malta, and the Netherlands were also included.

Sample size

The dataset covers 85 cohorts across seven countries. After removing duplicate attempts (retaining only first attempts) and applying inclusion criteria (a minimum of 12 completed pretest questions and 12 completed corresponding post-test questions, and a non-zero score on both pretest and in-course test), the final sample consisted of 899 students. This represents approximately 38% of the 2,390 students who originally enrolled across the 85 cohorts.

How much did participants learn?

Table 18 Mean results for the larger course population of the EBDM pre and in-course tests.

	Mean	SD
Pretest	40.8%	14.6
In-course post-test	62.8%	17.2
Difference	22.0%	18.8

Paired t-test: $t(898) = 34.99$, $p < .0001$ **Cohen's d = 1.17** (95% CI: [1.08, 1.25])

This is a **very strong result** - A Cohen's d of 1.17 is considered a very large effect, the 95% CI doesn't come close to zero, and the p-value is $p < .0001$. In plain English: students improved by an average of 22 percentage points from pretest to in-course test, a large effect that is unlikely to be due to chance and is considered educationally impactful.

Where did the most learning take place?

For details on the test results, see Appendix 11.7.

Note that the Cohens' d is lower at the question level. (The reason is that aggregation reduces variance while preserving the mean difference, which inflates Cohen's d at the total score level.)

Are the OzChild participants similar to the rest of the sample at baseline and follow-up?

Measure	OzC Mean (SD)	Rest Mean (SD)	t	p	Cohen's d	95% CI
Pretest (%)	45.1 (12.9)	40.7 (14.6)	1.55	.12	0.30	[-0.08, 0.69]
Posttest (%)	64.1 (16.0)	62.8 (17.2)	0.41	.68	0.08	[-0.30, 0.46]

The OzChild group showed a small difference in prior EBM knowledge compared to the rest of the sample ($d=0.30$, 4.4 percentage points higher on the pretest), but this is a small effect with wide confidence intervals given the size ($N=27$). At the in-course (post-test), the groups were virtually identical ($d=0.08$), a difference so small it is of no practical relevance regardless of sample size.

Pre-test and Post-test Scores by Student Group

The larger course population was also examined to determine if there was a difference in learning between those course participants with higher education and leadership levels. Table 19 shows the mean pre and post-test scores, indicating standard deviation, gain and Cohen's d values.

Table 19 Participants' pre and post-test scores by student group.

Group	N	Pretest M (SD)	Posttest M (SD)	Gain M (SD)	d	95% CI
Undergrad	516	40.0 (15.9)	62.7 (18.3)	22.8 (20.0)	1.14	[1.03, 1.25]
Master	263	41.9 (12.9)	62.5 (16.2)	20.6 (17.2)	1.19	[1.04, 1.35]
Executive	120	42.1 (12.1)	63.8 (14.2)	21.7 (17.0)	1.28	[1.04, 1.52]

Note. M = mean; SD = standard deviation; Gain = posttest minus pretest; d = Cohen's d for paired pre-post comparison; CI = confidence interval.

Table 20 shows the p values and $\eta^2 =$ eta-squared values that indicate no meaningful difference between the three groups. This indicates that EBDM training equally benefits participants irrespective of these educational and leadership levels.

Between-Group Differences: Undergrad, Master, and Executive Students

Table 20 The between group analysis showed no meaningful difference in test scores.

Outcome	F (2, 896)	p	η^2	Effect size	Interpretation
Pretest	2.04	.131	.005	Negligible	No meaningful difference between groups
Post-test	0.26	.773	.001	Negligible	No meaningful difference between groups

Outcome	F (2, 896)	p	η^2	Effect size	Interpretation
Gain	1.21	.299	.003	Negligible	No meaningful difference between groups

One-way ANOVA with $df = 2, 896$. $\eta^2 = \text{eta-squared}$ (proportion of variance explained by group membership). Benchmarks: .01 = small, .06 = medium, .14 = large (Cohen, 1988). All three η^2 values are well below the small threshold, indicating that student level (undergrad, master, executive) explains virtually none of the variance in pretest scores, posttest scores, or learning gains.

Limitations

- The outcomes are based on percentage scores that are bounded (0-100), which means they're not perfectly normally distributed — especially if many students score near 0 or 100. However, t-tests are fairly robust to this, especially with a large sample, such as indicated here.
- For individual question scores (Appendix 11.7), some questions are binary (0% or 100% correct), which makes t-tests less appropriate at the question level.
- However, for the overall pre/post scores (aggregated across all questions), the distribution will be much more normal and t-tests are perfectly appropriate.
- Students who score below 25% correct on the post-test most likely did not engage (i.e., did not complete the coursework and guessed on their first attempt). For this reason, it may be wise to run a sensitivity analysis excluding these students.
- Regarding the comparison of OzChild participants with the rest of the sample: a sample size of $N=27$ is rather small for this kind of comparison, which means the test has limited statistical power to detect moderate differences (if they would exist.) However, a Cohen's d of .08 is so close to zero that even with a much larger OzChild sample it would almost certainly remain negligible.

Data Visualisations

Statistical significance is one way to derive meaning from data. Another insight is to examine the way thinking changes between participants prior to the intervention (in this case the EBDM programme) and afterwards. One way to do this is to create a chord diagram, which in this case illustrates relationships between codes *within* an uttering. Because this study employed multi-label coding, allowing more than one code to be chosen in an utterance, relationships between codes can be insightful. In this way insights can be gained into the changes in thinking between the two "think aloud" sessions. For the chord diagrams, Flourish Studio was used (Canva, 2026).

Chord diagrams show the codes on the outside, with connections between codes in the middle. The larger the area, the more mentions of that particular code. If a code is not mentioned by itself (without connections), it does not appear. What is insightful are the connections between codes.



Figure 9 Pre intervention think aloud for the course group overall.

For the post-intervention course group, those who completed the Certificate in Evidence-Based Decision-Making for Human Service Leaders programme, the word cloud is more nuanced. The word "around" features highly, with "think", "information", "looking" and "problem" sharing equal frequencies. This could indicate that the participants were examining the problem more extensively and seeking information about it.



Figure 10 Post intervention "think aloud" Control group overall.

Figure 12 Participant 4 Pre-intervention "think aloud" Chord Diagram

Figure 12 shows very few "Scientific data" mentions (purple) and few "Search" mentions (light orange). Similarly, there are mentions of practitioners, but only in the context of the "Appraising" and "Aggregating" codes, indicated by the lines with dark green at the tip.

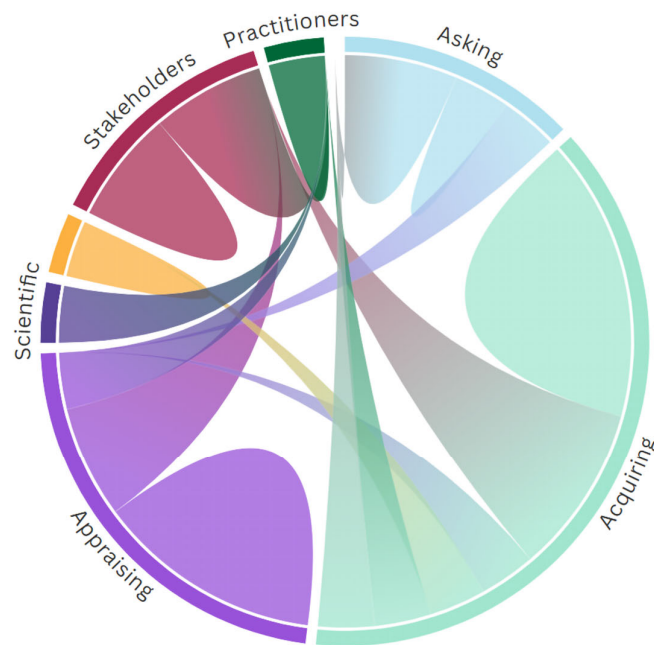


Figure 13 Participant 4 Post-intervention "think aloud" Chord Diagram

Figure 13, the post-intervention chord diagram shows connections between "Acquiring", "Stakeholders" and "Practitioners". Also, connections between "Scientific data" and "Practitioners". There are also connections between "Asking" and "Scientific data". "Appraising" is occurring in the context of the "Practitioners".

When comparing the two chord diagrams, (Figure 12, Figure 13) one can see that there are more utterances in the "Scientific data" and "Practitioners" codes in the post-intervention "think aloud" data. There is more thought into "Acquiring" data from "Stakeholders" than in the first "think aloud" exercise. There are fewer utterances in the "Acquiring" code in the second "think aloud" session, and there are no utterances evident in the "Culture" code. The second "think aloud" session showed more "Asking" than the first, but the pre-intervention "think aloud" showed more "Search".

In the post-intervention word cloud the participant uses the words "professionals" (which is in the practitioner's code) and "understand" and "understanding" more frequently.

"The professionals, so I'm assuming we are referring to healthcare professionals. I would be engaging with them through perhaps a couple of different avenues around how I can gain some further understanding."

Interestingly the word "information" has equal prominence in both "think aloud" sessions. The words "around" and "probably" also dominate, but are similar in frequency in both, indicating that this might be words of preference for this participant.

In the quantitative test results for Participant 4, initially the score was 36%. The in-course tests showed an improvement, which rose to 81%. The post-test was completed some months later, which returned a result of 54%. This indicated an improvement in score of 18% between pre- and post-test results. The drop from the in-course test and the post-test is called the Forgetting Curve. Without reinforcement memory of course materials fade (Murre & Dros, 2015).

Participant P8 - Course Cohort

Participant 8 (P8) was a member of the Certificate in Evidence-Based Decision-Making for Human Service Leaders (course cohort).

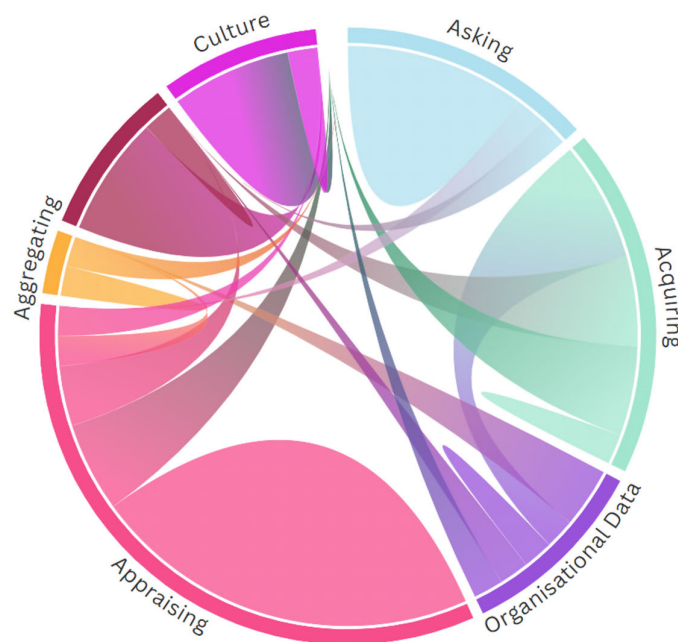


Figure 16 Participant 8 pre-intervention "think aloud" chord diagram (Canva, 2026)

Figure 16 shows the pre-intervention "think aloud" data for Participant 8, with connections occurring between "Organisational Data" and "Aggregating". There is also "Appraising" of "Scientific data" which is also mentioned in the context of "Organisational data" but is not prominent in itself. There are some connections between "Acquiring" and "Organisational data" as well as "Practitioners" and "Stakeholders".

"Then it's really about say getting these managers together so that the CFO, your HR executive group and other seeking evidence I suppose, or feedback. And that can be done via surveys or interviews In terms of one-on-one chats with people, leaders with the actual staff themselves, it's obviously the key stakeholder, the people that we can ask the question around with that demographic, the age group around that we're actually targeting to understand what their needs and wants and desires are and whether they're looking to stay or move or what would help retain them and actually try to narrow down and understand their motivations and solutions based off that."

However, "Practitioners" do not feature heavily in the thinking. There is some mention of "Aggregating" "Organisational data" also, but this does not heavily feature, as the lines are quite thin.

"There's a lot of assumptions there. We haven't seen the evidence based off our HR data and what's actually out there in society and similar companies as well."

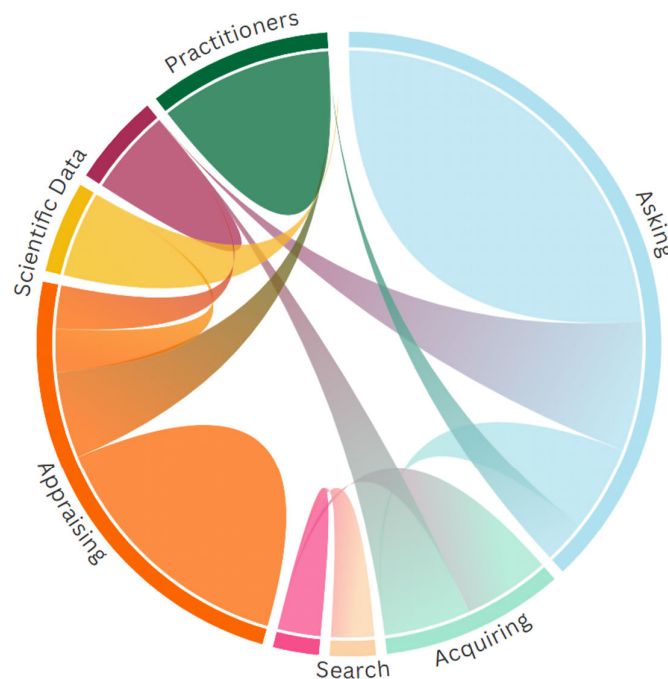


Figure 17 Participant 8 post-intervention "think aloud" chord diagram. (Canva, 2026)

Figure 17 shows more activity in the "Asking" code, with relationships between it and "Stakeholders" and "Practitioners". "Acquiring" occurs less frequently, with connections between it and "Stakeholders". There is some "Appraising" of "Practitioners".

Figure 19 Participant 8 post-intervention "think aloud" word cloud (Freewordcloudgenerator, 2026)

The word clouds for Participant 8 (Figure 18, Figure 19) show a marked increase of the word "evidence" in the post-intervention thinking:

"Who agreed on the targets and what evidence was used to actually define those targets? And you've got some different types of, I suppose, evidence there. But going back to the different sort of the four pillars of evidence, you've got practitioners, evidence from the organization, scientific literature, and which I missing? And stakeholders, I think it was, we've got some assumptions here." (Post intervention "think aloud")

This sort of thinking aligns closely with the tenets of evidence-based decision-making.

Participant 8 in the pre-intervention "think aloud" mentions the problem and the solution, as well as mentioning "people" frequently.

Whereas in the post-intervention "think aloud" the focus is on the "targets" (see Appendix 11.3 "Think aloud" scenarios) and the "assumptions" behind them. There is also slightly more emphasis on stakeholders (patients, directors), but when balanced with "employees" in the pre-intervention "think aloud" the differences in the practitioner's code does not appear to be great.

In the quantitative results, P8 did not complete the post-test. The pre-test mark was 47% and the in-course mark was 81%. This was a difference of 18 marks. This aligns with the larger course population that showed a mean improvement of 22%.

Participant 10 - Course Cohort

Participant 10 (P10) was a member of the Certificate in Evidence-Based Decision-Making for Human Service Leaders (course cohort).

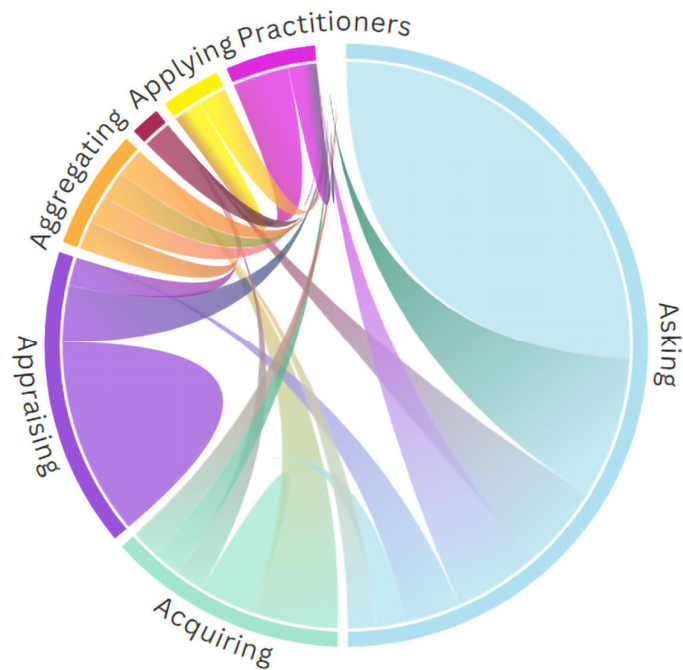


Figure 20 Participant 10 pre-intervention "Think aloud" chord diagram

In the Figure 20 chord diagram linkages between "Acquiring", "Asking" and "Practitioners" is evident, however "Practitioners" does not feature by itself solely (lines becoming green). There are linkages between "Organisational data" and "Stakeholders" as well as "Asking" and "Organisational data".

"What is going on with other teams and team leaders or management in that kind of with similar or the same agency with different teams so.

Learning about what they do and how they kind of tackle the situation that will be helpful. I guess I'll go with them and talk about that situation and go through if they had similar experience and what was their approach and what did work for them and like exploring options around."

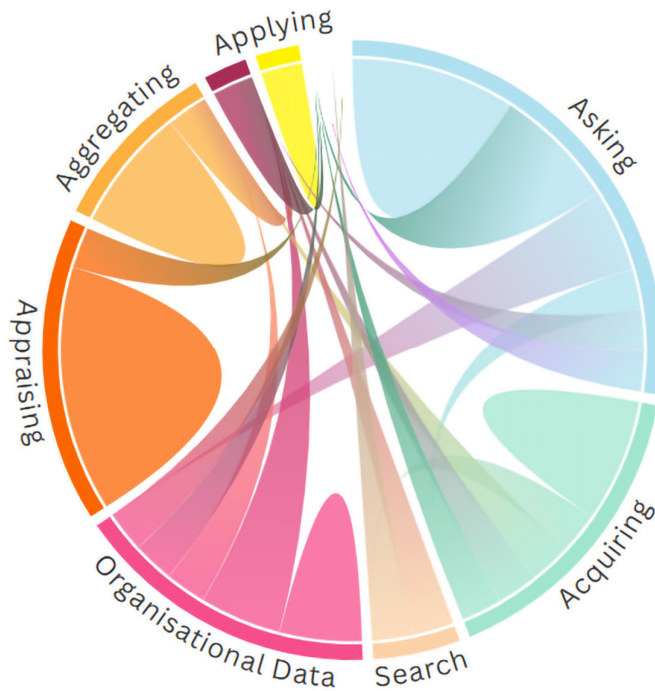


Figure 21 Participant 10 post-intervention "think aloud" chord diagram.

In both word clouds, the word "kind" appears quite frequently. This participant appears to use the phrase "kind of" as a part of their personal lexicon, which would explain its dominance. In the pre-intervention word cloud (Figure 22), the words "situation" and "team" as well as "happening" occur frequently, but overall the word cloud is quite spare in repetitions.

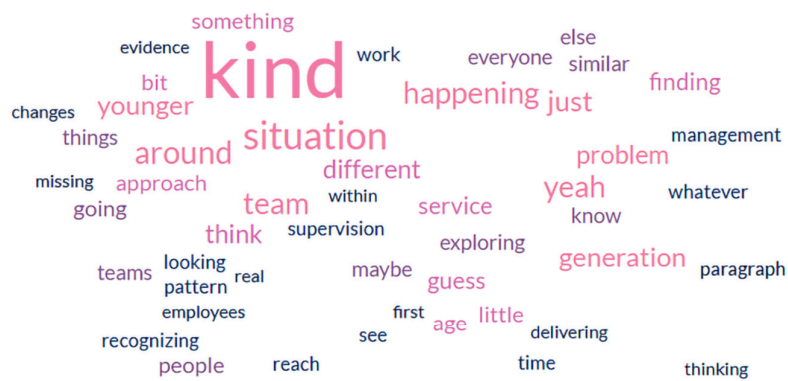


Figure 22 Participant 10 pre intervention "think aloud" wordcloud course cohort

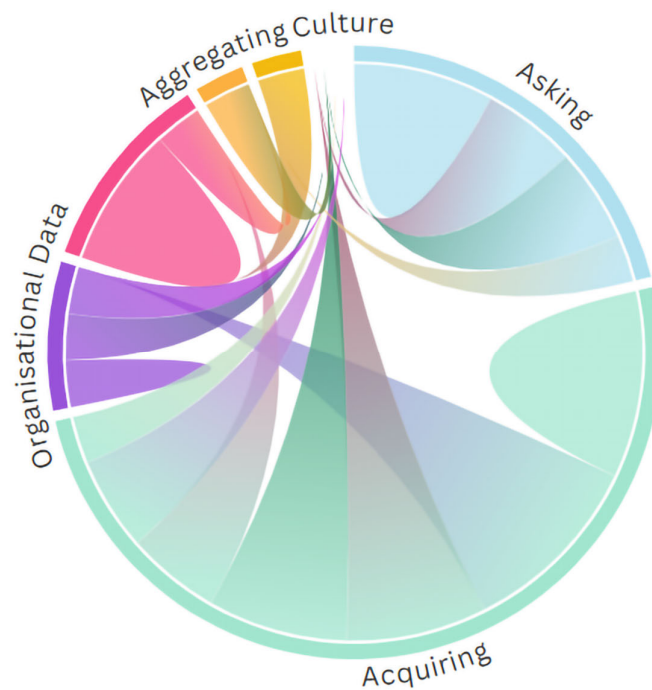


Figure 24 Participant 12 pre-intervention "think aloud" chord diagram

In the pre-intervention "think aloud" data, Participant 12 (P12) did not mention "Scientific Data" at all. "Search" was only mentioned once in relation to "Acquiring". The two codes with the most connections were "Acquiring" and "Organisational Data". Similarly, "Acquiring" in relation to "Stakeholders" and "Practitioners" show equal frequencies. "Culture" was not mentioned by itself but was in relation to "Acquiring" and "Organisational Data".

"Get a little bit of evidence, I guess, around why their employees are leaving rather than making assumptions about that. I guess the assumed problem is that their employees are leaving because of the team leaders, but there's no evidence to back that up. So, I guess I'd go back to exit data, seeing if there's anything that we can find any common threads from what's already been given to us, making sure that we've got interviews included in the process and then everybody can benefit from understanding, including the CEO, that there are generational differences."

In the post-intervention "think aloud" data, P12 had a low number of utterings. This participant had a very short session, with only 13 utterings that could be coded. This makes the data significantly less reliable than other datasets. The participant admitted difficulty in remembering the course learnings.

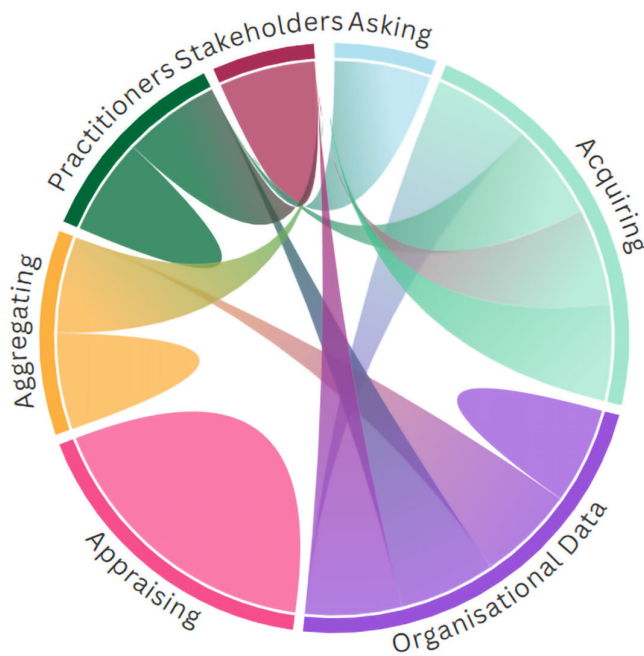


Figure 25 Participant 12 post-intervention chord diagram is unreliable due to less data.

However, despite the small number of utterings, the participant did appear to consider the "Stakeholder" and "Practitioner" codes more frequently. There was a desire to consider these codes in conjunction with "Organisational data" and "Acquiring". "Scientific data" was considered once, in conjunction with "Aggregating".



Figure 26 Participant 12 pre-intervention "think aloud" word cloud

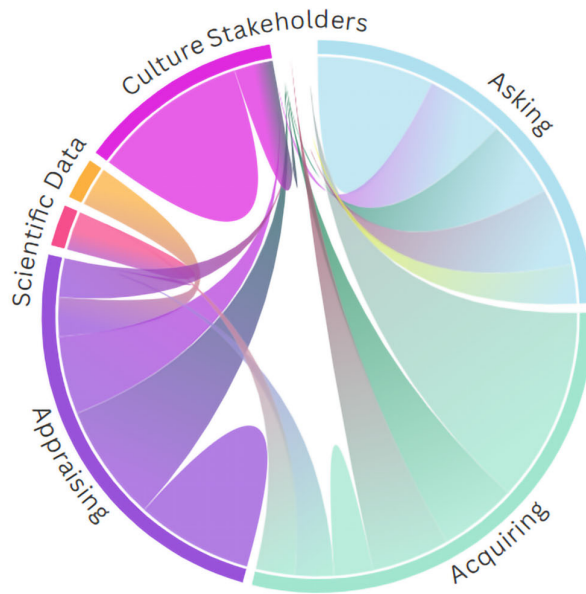


Figure 28 Participant 14 pre intervention "think aloud" chord diagram

Again, when examining Participant 14's (p14) before and after intervention chord diagrams (Figure 28, Figure 29), one can clearly see more mentions of "Stakeholders" in the second "think aloud" session. In the first diagram (Figure 28) "Practitioners" are clearly mentioned in association with "Acquiring", as are "Stakeholders" but they are never mentioned alone. "Practitioners" are also mentioned in relation to "Appraising". Culture features more than usual, mentioned by itself and in relation to "Asking".

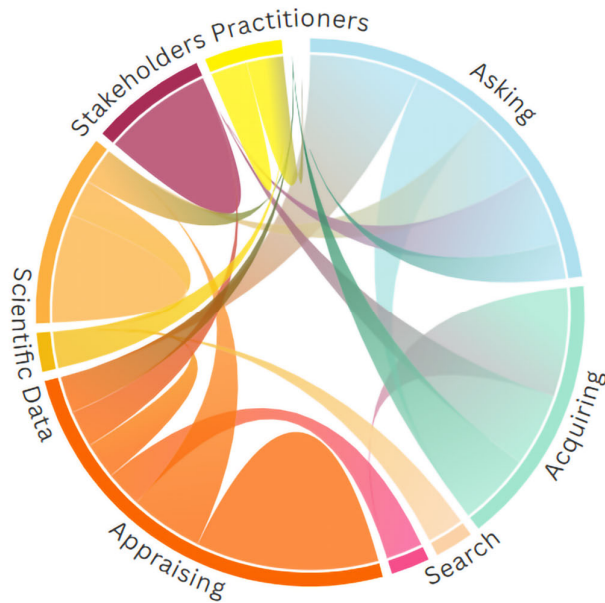


Figure 29 Participant 14 post-intervention "think aloud" chord diagram

In the post-intervention "think aloud" chord diagram, the picture is more complex. There are more diverse intersections overall, with "Search" making an appearance. There appears to be more "Aggregating" (light orange) occurring, mentioned by itself and in relation to "Asking" "Scientific Data" and "Practitioners". "Stakeholders" are mentioned on their own three times (maroon) and in relation to "Asking" and "Acquiring", indicating more thought being put into asking stakeholder's opinions and acquiring data on them. There are also mentions of "Acquiring" and "Asking" of "Practitioners", although they are not mentioned on their own.

"So, I would be saying that the practitioner evidence here is really important as is actually some data around, they're actually saying there as well that it could be an administrative problem. So, I would be seeking some data actually in the first and foremost, I'd be looking at the organizational data." (Participant 14 post "think aloud data")

There are three instances of "Acquiring" "Organisational data" in the utterances.



Figure 30 Participant 14 pre-intervention "think aloud" word cloud



Figure 31 Participant 14 post-intervention "think aloud" word cloud

Participant 14 did not show statistically significant differences between the pre- and post-intervention "think aloud" data. This is reflected in the word clouds, where many of the same phrases are employed. The use of "Looking" and "around" appears to be part of the participant's personal lexicon. The participant considered "information" more frequently in the pre-intervention "think aloud". Other differences appear to reflect the different scenarios, with a slight increase of "scientific" in the post-intervention word cloud (Figure 31).

In the quantitative results, Participant 14 scored 42% on the pre-test and 94% on the post test. This was an improvement on their course mark of 83%. This is because this participant omitted to complete one module's assessment in the course.

Participant 1 - Control Cohort

Participant 1 (P1) was NOT a member of the Certificate in Evidence-Based Decision-Making for Human Service Leaders (course cohort). This participant was a senior leader in the OzChild organisation with no prior training in EBDM. They were a member of the Control cohort.

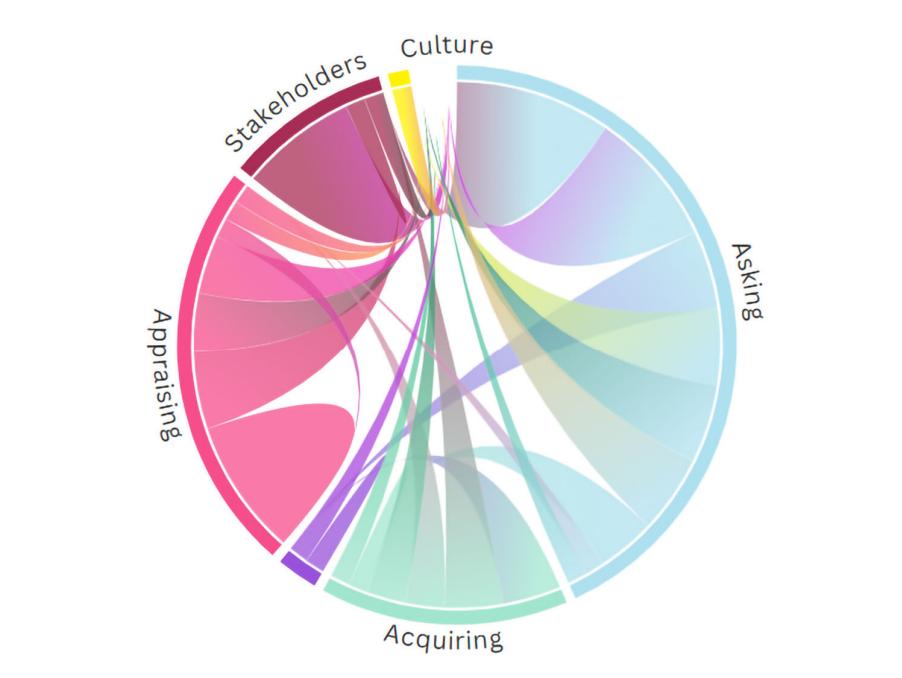


Figure 32 Participant 1 pre-intervention "think aloud" chord diagram

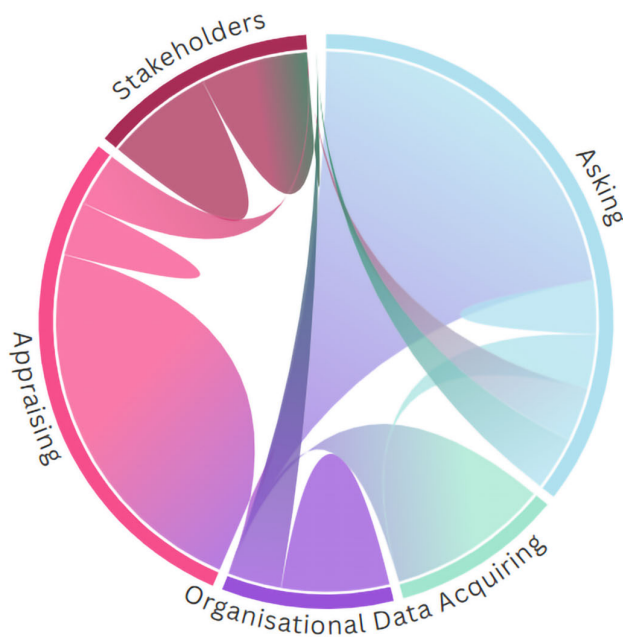


Figure 33 Participant 1 post-intervention "think aloud" chord diagram.

This participant did show a statistically significant difference between the two "think aloud" sessions. The pre-intervention data contained 54 utterings. The post-intervention "think



Figure 35 Participant 1 post-intervention word cloud.

The differences in the word clouds reflect the differences in the two scenarios, with "people" emphasised in the first and "targets" and "time" in the second. The first scenario was about people retention, the second about missing targets (see Appendix 11.3). Figure 34 shows more mention of "management", Figure 35 shows more mentions of "organization" and "patients".

In the quantitative results, P1 scored 43% in the pre-test results and 39% in the post-test results. This showed a decrease of 4%.

Participant 3 - Control Cohort

Participant 3 (P3) was a member of the Control cohort. This participant was a member of senior leadership in OzChild with no previous training in EDBM.

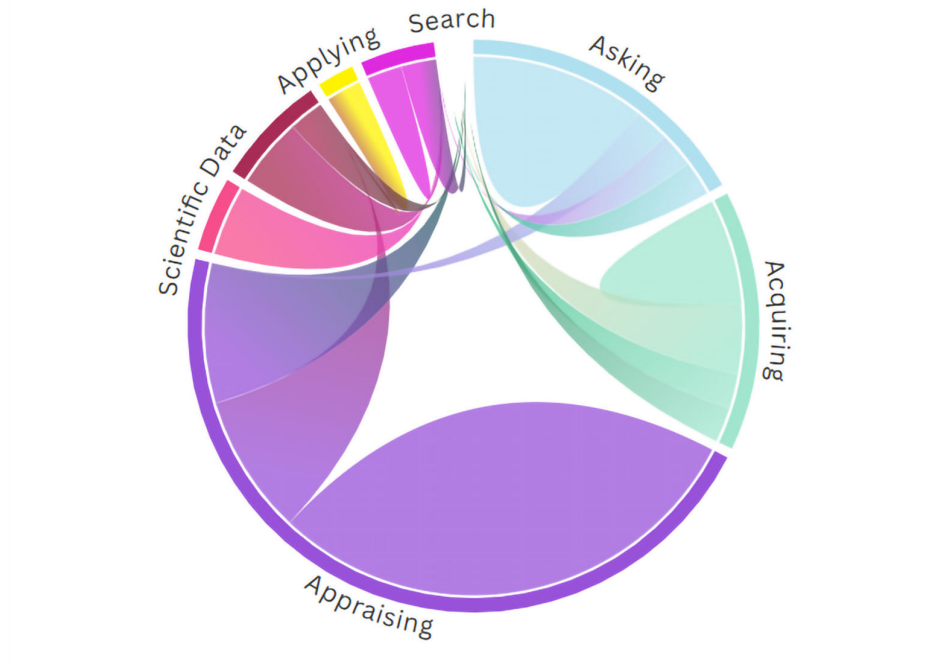


Figure 36 Participant 3 pre-intervention "think aloud" chord diagram.

Participant 3 spent time "Appraising" "Stakeholders" and "Participants" (Figure 36). More than most participants, there was consideration between "Scientific data" and "Culture" (on 2 occasions). There were no utterances on "Assessing" the data, nor on "Aggregating" the data, comparing it with external companies.

By comparison Figure 37 shows more focus on "Acquiring" "Organisational Data". This participant focussed heavily on "Organisational data", in terms of "Acquiring" and by discussing it solely. There was no mention of "Scientific data" or "Culture" in this data. There is much less focus on "Appraising" in the second session. There is no mention of "Assessing" in this data.

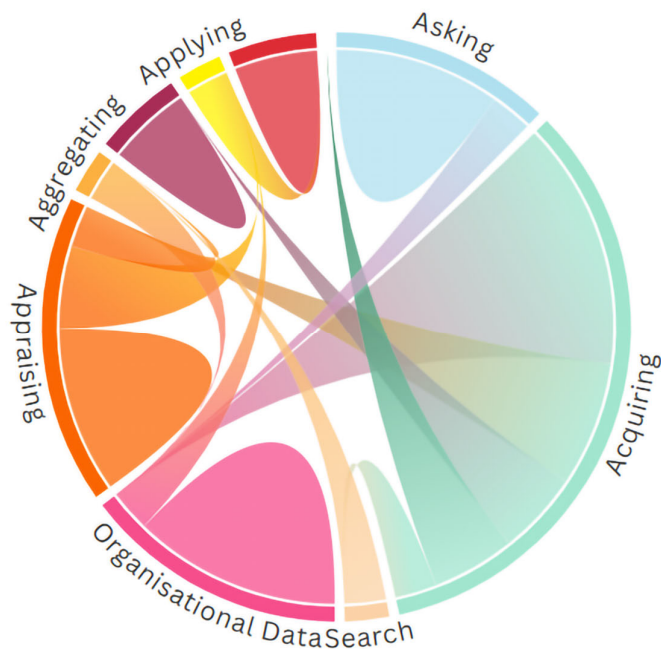


Figure 37 Participant 3 post-intervention "think aloud" chord diagram.



Figure 38 Participant 3 pre-intervention "think aloud" word cloud.

Figure 38 shows consideration of "people", "management" and "statistics". Also "evidence", "workforce" and "personal" are mentioned frequently. The first scenario deals with young people leaving a company, so the emphasis on "people" is understandable.

“Yeah, so there's an aging workforce, so they probably need to do something around increasing younger people into the company or bringing younger people into the company. Again, there's plenty of statistics out there that would, or management styles that would be able to be implemented that could help that.”

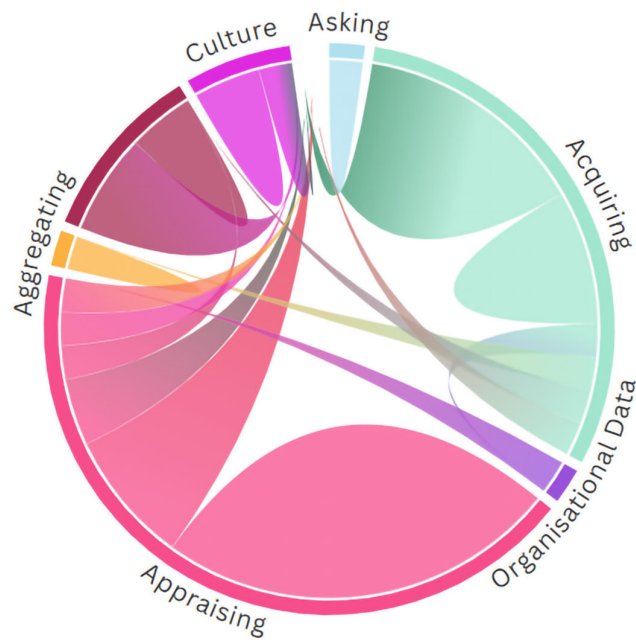


Figure 40 Participant 5 pre-intervention "think aloud" chord diagram

In the pre-intervention "think aloud" chord diagram (Figure 41), Participant 5 did not use the "Search" or "Assessing" codes. There is only one mention of "Organisational data" (purple) in the context of the "Appraising" code. Practitioners do not feature by themselves but frequently occur in conjunction with the "Acquiring" code (7 times). This participant "Appraised" "Scientific data" 4 times (largest red line from Appraising), which does appear to be more frequent than the norm.

"Culture" is mentioned a few times, twice by itself and once in conjunction with "Practitioners".

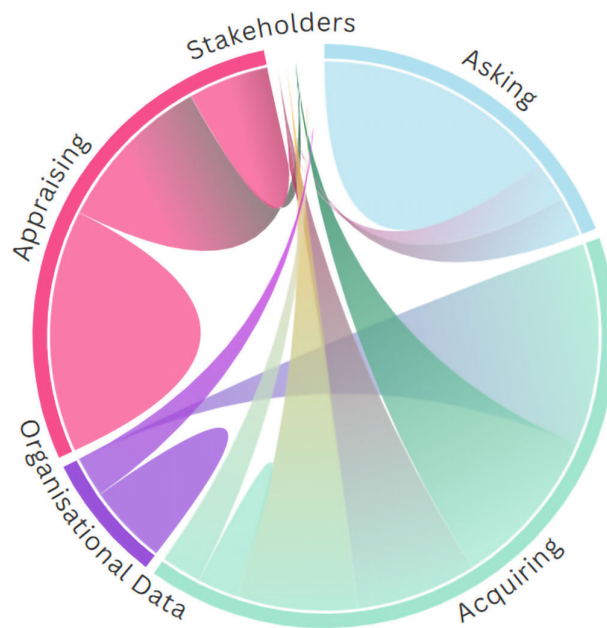


Figure 41 Participant 5 post-intervention "think aloud".

In the post-intervention "think aloud" for this Control group participant, there is "Acquiring" of "Organisational data" (5 times) and "Acquiring" mentioned with "Practitioners", (4 times). This is less than the pre-intervention "think aloud". There is no mention of "Scientific data", "Applying" or "Assessing". Culture is only mentioned once, in the context of "Organisational data". Overall, the two chord diagrams look similar.

In Figure 42, the participant does appear to consider the "evidence" in this scenario. P5 refers to "research" and "approach" frequently.

"If I had this problem, I would be looking at some of the HR research or the research that's gone into millennials and Gen Zs, whether there's any evidence around that. They respond better to different coaching style of management than to a more traditional top-down authority approach. In addition, they prefer a more collaborative in person way. Yeah, I'd just be looking for more evidence around some of this decision making that's going on here."

In the quantitative results, P5 had a pre-test score of 45% and a post-test score of 47%. This showed an increase of 2% and cannot be considered significant.

Participant 7 - Control Cohort

Participant 7 (P7) was a member of the Control cohort. This participant was a member of senior leadership in OzChild with no previous training in EDBM.

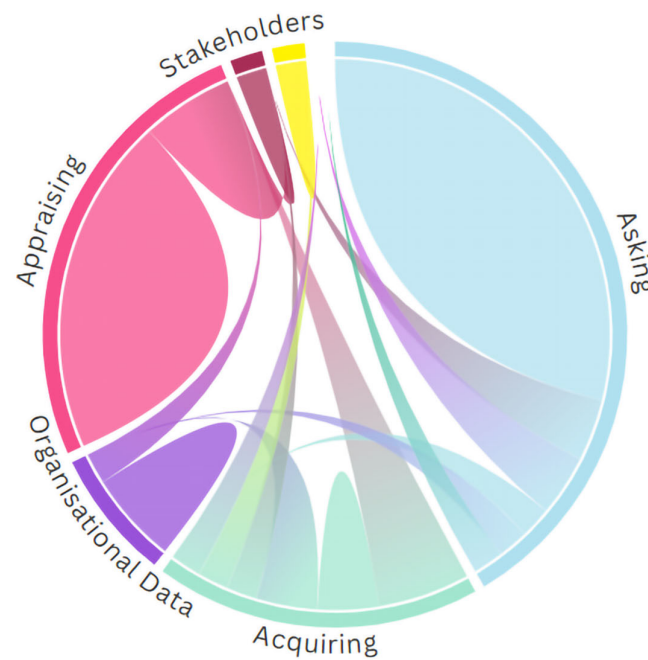


Figure 44 Participant 7 pre-intervention "think aloud" chord diagram

Participant 7 interrogated scenario one with more questions than usual. However, there was no mention of "Scientific data", "Aggregating" or "Practitioners". There was some mention of "Stakeholders" and "Appraising" "Stakeholders". This participant was interested in "Acquiring" information from many sources, as indicated by the lines emerging from that area.



Figure 47 Participant 7 post-intervention "think aloud" word cloud.

Figure 46 and Figure 47 also show a difference in the number of different words used between the two scenarios. The pre-intervention word cloud appears more populated. There is more consideration of "evidence" in the pre-intervention "think aloud" word cloud than the post-intervention data. The emphasis on "people", "patient" and "therapist" does show consideration of stakeholders and practitioners. The main differences appear to be scenario-based, rather than the tenets of EBDM.

"I just guess having a, looking at from trying that evidence-based approach, is it looking at mapping all the therapists, the types of treatment that people, what they're needing are some more than others? Is it taking more, like if somebody with an eating disorder, is that a longer a time period than maybe somebody with an anxiety disorder?" (Participant 7 post-intervention "think aloud").

The quote above is indicative of the large number of questions (Asking) in the post-intervention "think aloud" for this participant.

In the quantitative results, P7 had a pre-test score of 37% and a post-test score of 57%. This showed an increase of 20%. Although this does indicate social learning, a great improvement on the pre-intervention test results, the word cloud does not indicate a change in overall thinking.

Participant 9 - Control Cohort

Participant 9 (P9) was a member of the Control cohort. This participant was a member of senior leadership in OzChild with no previous training in EDBM.

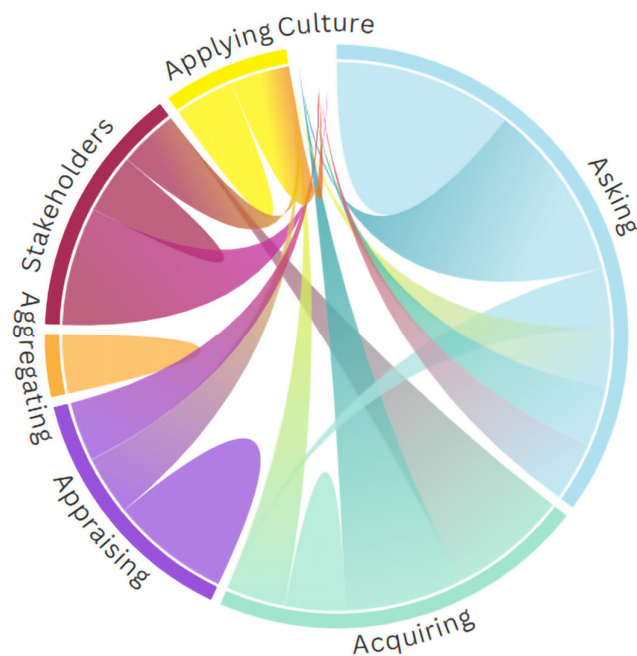


Figure 48 Participant 9 pre-intervention chord diagram.

Figure 48 shows "Acquiring" from "Stakeholders" and "Organisational data". Interestingly this participant uses "Applying" more frequently than other participants, essentially agreeing with the proposed solution in the scenario.

"Yeah, I think the solution would still be the same that you know, there is obviously evident there's a problem that they've identified, so it requires a solution, but that would be one solution I would think about". (Participant 9 pre-intervention "think aloud").

There are also links between "Stakeholders" and "Culture" as well as "Asking" "Stakeholders" and "Organisational data." This participant discusses interventions with managers, however, does not mention employees (Practitioners code) at all. Scientific data is also absent from this "think aloud", as is "Search".

larger course population. All participants showed an increase in test scores of 22%, which is statistically significant.

When the larger EBDM course population is examined by student group, in terms of Undergraduate, Masters or Executive level participants, there is no statistical significance between the pre and post-test results. This appears to indicate that EBDM training is equally beneficial to participants notwithstanding educational or leadership levels.

Looking at the results in the pre-intervention "think aloud" sessions between the smaller OzChild EBDM course group and the Control group, there are differences in the "Asking" and "Acquiring" areas (see Figure 4), and a small difference in the "Scientific data" area.

Comparing the OzChild Control group with the OzChild EBDM course cohort (see Figure 6), for the post-intervention results, there is more emphasis in the EBDM course cohort in "Appraising" as well as "Stakeholders" and "Practitioners", with less emphasis on "Organisational data", which featured strongly within the Control group. The Control group indicated a lack of reference to "Scientific data", whereas they considered this more frequently in the pre-intervention "think aloud". This shows marked differences between the two groups.

This indicates an increase in evaluating the data from stakeholders and practitioners in the OzChild EBDM course cohort, and there was increased consideration of scientific data.

The Control group findings are equally instructive. The continued reliance on organisational data and the absence of scientific evidence post-intervention reinforce what we know from the literature: without a structured intervention, leaders default to experience and internal data.

There was also a reluctance on the part of the EBDM course group to move to the "Apply" stage in the post-intervention "think aloud" data, which indicates a preference for evaluating the data before moving to a preferred solution. The participants from the OzChild EBDM course appeared to place greater emphasis on examining the evidence and did not want to be drawn on conclusions or possible solutions before scrutinizing the data.

This is the most significant finding, this shift in leaders' cognitive approach to decision-making. The trained group is clearly engaging in more deliberate appraisal, drawing on broader evidence sources and resisting premature conclusions. This is an important shift and aligns with expected changes when exposed to EBDM principles.

This was reflected in the individual data, which appeared to indicate the same results. The chorded diagrams gave insights into the connections *within utterings* to indicate the nuances in the thinking, given this is multi-label data.

The word cloud data when viewed as a whole, indicated more frequent utterings of key EBDM codes by the OzChild EBDM course cohort, when compared with the OzChild Control group, which also indicated fewer repetitions.

The main differences in the two "think aloud" sessions in the OzChild Control group appeared to rest in the differences between the two scenarios, rather than an overall difference in approach or thinking.

In both groups, for both "think aloud" sessions, there is negligible mention of the "Assessing" code. The absence of the "Assessing" phase across both groups is particularly important. From both a research and practice perspective, this highlights a systemic gap in evaluation and learning that is critical in human services, where outcomes and accountability matter. This represents a gap in the research process, where the application of the "Assessing" phase was not possible using fictional scenarios. The evaluation of the use of the "Assessing" phase would require further research embedded within practice.

The quantitative data from the pre- and post-intervention tests consistently showed an increase of the scores from the OzChild EBDM course cohort. The Control group showed improvements in some participants, with about half having negligible results. This data is unreliable due to the small sample. The improvement in learning is reflected in the quantitative results from the larger course population, at the top of this section.

However, the forgetting curve is evident, even given the small sample. Without reinforcement and application, capability diminishes quickly. Leadership capability is necessary but not sufficient for implementation. Organisational climate, systems and reinforcement mechanisms are critical in determining whether evidence-based approaches are sustained in practice. The use of EBDM principles becomes an organisational capability. It requires shared language, reinforcement, and leadership modelling to be translated into practice. Without clear structures, EBDM principles are used for analysis, but are not necessarily translated in practice.

7. Limitations

One of the limitations in this study was the small number of participants in the OzChild EBDM course cohort and Control group. Originally there were 14 participants, from an initial goal of 20 (10 in control, 10 in course cohort). As this was a longitudinal study, two participants discontinued with the EBDM programme, one resigned from their position and to keep the numbers even, one was excluded from the programme.

This left an overall sample size of 10 participants - five from the control and five from the course cohort. Although the sample was small, it was possible to obtain statistical information on the qualitative data. However, for the test results, the sample was too small to show any significant results, hence the need for a comparison with the larger course cohort, who were not from the OzChild organisation. Therefore, any reliance on statistical significance on the test results from the OzChild EBDM course cohort and Control group should be treated with caution.

8. Conclusions

To answer research question one:

RQ1: *To what extent does the Certificate in Evidence-Based Decision-Making for Human Service Leaders produce changes in the application of evidence-based principles and practices in decision-making among executive leaders in the child welfare sector?*

We see more mentions from the EDBM course group in the codes "Appraising", "Scientific Data", "Stakeholders" and "Practitioners" (Figure 6, Figure 7). In the Control group, "Scientific data" is absent, but there is more reliance on "Organisational data". There is a statistically significant difference between the pre- and post-intervention data for the EDBM course group when viewed as a whole. There is no statistical significance for the Control group.

Research question 1a asks:

RQ1a: *What are the specific changes in decision-making behaviours and strategies observed among leaders in the child welfare sector following the completion of the Evidence-Based Decision-Making for Human Services Leaders program?*

There is more reliance on "Acquiring" information from "Stakeholders" and "Practitioners" in the EDBM course group, as seen in the chord diagrams. There also appears to be more consideration of the EDBM codes seen in the word clouds, with the Control group showing less repetition overall. The main differences in the Control group appear to reflect the differences in the scenarios, rather than differences in thinking.

The quantitative results appear to indicate an overall increase in results from the EDBM course group, but the Forgetting Curve moderates this effect over time.

The Control group showed half of the participants showing an increase in results with half with negligible differences. However, the sample size is too small to draw conclusive results.

Research question 1b asks:

RQ1b: *What are the differences, if any, in the change in decision-making quality between the leaders who participated in the Certificate in Evidence-Based Decision-Making for Human Services Leaders program and a control group of leaders who did not receive the program?*

There is more consideration of stakeholders and practitioners in the EDBM course group, and less emphasis on organisational data. There is a marked increase in the use of scientific data and an absence of that in the Control group. The EDBM course cohort appears to refrain from settling on a solution, with less use of the "Applying" and "Assessing" codes. This would appear to indicate the desire to gather data and evaluate that rather than deciding on a solution at an early stage.

9. Author Contributions:

Fiona M. Chatter: Conceptualisation (Lead role - Ethics application), Methodology, Validation, Lead role data collection, Lead role Formal Analysis, Investigation, Data Curation, Writing (Original Draft & Editing), Visualisation.

Roslyn Cameron: Conceptualisation, (Supporting role - Ethics application & Submission) Validation, Methodology, Supporting role - Data collection, Software - Transcriptions (REV.com), Supporting role - Formal Analysis, Writing (Review & Editing), Supervision, Project Administration, Funding Acquisition.

Lisa J. Griffiths: Conceptualisation, Writing (Original Draft - Ethics application, Literature review), Funding Provision, Resources.

Eric Barends: Conceptualisation, Resources (Course content, scenarios, pre- and post- tests), Supporting role Formal Analysis.

Timothy Lippis: Conceptualisation, Resources and PAC Coordination.

Jarrod Edwards: Conceptualisation, Supporting role - Project Administration.

Denise M. Rousseau: Conceptualisation, Resources.

10. References

- Aarons, G. A. (2006). Transformational and transactional leadership: Association with attitudes towards evidence-based practice. *Psychiatric Services*, 57(8), 1162–1169.
- Aarons, G. A. , & Sawitzky, A. C. (2006). Organizational culture and climate and mental health provider attitudes toward evidence-based practice. *Psychological Service*, 3(1), 61.
- Álvarez-Correa, A. A., & Quintero-Arrubla, S. R. (n.d.). *Meaningful learning: the end of the social and individual constructivist dichotomy? Aprendizaje significativo: ¿el fin de la dicotomía social e individual constructivista?*
- Australian Institute of Family Studies. (2017). *Child protection and Aboriginal and Torres Strait Islander children*. Australian Government. <https://aifs.gov.au/resources/policy-and-practice-papers/child-protection-and-aboriginal-and-torres-strait-islander>
- Australian Institute of Family Studies. (2018). *The economic costs of child abuse and neglect*. Australian Government. <https://aifs.gov.au/resources/policy-and-practice-papers/economic-costs-child-abuse-and-neglect>
- Barends, E., Villaneueva, J., Rousseau, D. M., Briner, R. B., Jepsen, D. M., & Houghton, E. (2017). Managerial attitudes and perceived barriers regarding evidence-based practice: An international survey. *Plos One*, 12(10).
- Bish, A., & Becker, K. (2016). Exploring expectations of nonprofit management capabilities. *Nonprofit and Voluntary Sector Quarterly*, 45(3), 437–457.

- Burgemeister, F. C., Hokke, S., Crawford, S. B., Hackworth, N. J., & Nicholson, J. M. (2022). Does place matter in the implementation of an evidence-based program policy in an Australian place-based initiative for children. *Health & Social Care in the Community*, 30, e5786–e5800.
- Canva. (2026, March 12). *Flourish Studio*. Canva. <https://app.flourish.studio/>
- Capili, B., & Anastasi, J. K. (2024). An Introduction to the Quasi-Experimental Design (Nonrandomized Design). *American Journal of Nursing*, 124(11), 50–52.
- Carnochan, S., McBeath, B., & Austin, M. J. (2017). Managerial and frontline perspectives on the process of evidence-informed practice within human service organizations. . *Human Service Organizations: Management, Leadership & Governance*, 41(4), 346–358.
- Corrigan, P. W., Lickey, S. E., Champion, J., & Rashid, F. (2000). Mental health team leadership and consumers' satisfaction and quality of life. *Psychiatric Services*, 51(6), 781785.
- Covey, S. R. (2004). *The 7 habits of highly effective peop* (7th Edition). Simon & Schuster.
- Creswell, J. W. (2013). *Qualitative Inquiry and Research Design: Choosing among Five Approaches*, (3rd ed.). Sage.
- Dawson, C., Julku, H., Pihlajamäki, M., Kaakinen, J. K., Schooler, J. W., & Simola, J. (2024). Evidence-based scientific thinking and decision-making in everyday life. *Cognitive Research: Principles and Implications*, 9(1), 50. <https://doi.org/10.1186/s41235-024-00578-2>
- Department of Family and Community Services (FACS). (2016, October 23). *NSW Budget - reforms for kids needing care*. NSW Government.
- Department of Health and Human Services. (2016). *Roadmap for Reform: Strong Families, Safe Children. The first steps* . Victorian Government .
- Doolittle, P. E., & Hicks, D. (2003). *Theory and Research in Social Education Winter* (Vol. 31, Number 1).
- Finn, N. K., Torres, E. M., Erhart, M. G., Roesch, S. C., & Aarons, G. A. (2016). Cross-validation of the Implementation Leadership Scale (ILS) in child welfare service organisations. *Child Maltreatment*, 21(3), 250–255.
- Free Word Cloud Generator. (2026, March 13). *Free Word Cloud Generator*. <https://www.freewordcloudgenerator.com/>
- Glisson, C., & Green, P. (2011). Organizational climate, services, and outcomes in child welfare systems. . *Child Abuse & Neglect*, , 35(8), 582–591.
- Glisson, C., & Hemmelgarn, A. (1998). The effects of organizational climate and interorganizational coordination on the quality and outcomes of children's service systems. *Child Abuse & Neglect*, 22(5), 401–421.
- Goldman, H. H., & Azrin, S. T. (2003). Public policy and evidence-based practice. *Psychiatric Clinics*, 26(4), 899–917.

- HakemZadeh, F., & Rousseau, D. M. (2024). Evidence-based decision-making is a social endeavor. *Behavioral Science & Policy*, 10(1), 27–33.
- Kaplan, M. (2025, May 19). *How to Use Krippendorff's Alpha to Measure Annotation Agreement*. Label Studio. <https://labelstud.io/blog/how-to-use-krippendorff-s-alpha-to-measure-annotation-agre>
- Kaushik, V., & Walsh, C. A. (2019). Pragmatism as a research paradigm and its implications for Social Work research. *Social Sciences*, 8(9). <https://doi.org/10.3390/socsci8090255>
- Kibort, P. M. (2004). The written word: Literature and leadership. *Physician Executive*, 30(4), 10.
- Krippendorff, K. (2011). *Reliability Data Observed Difference a-Agreement*. <https://repository.upenn.edu/entities/publication/034a6030-c584-4d14-9d3d-7b7e8d16df20>
- Lansford, M., Clements, V., Falzon, T., Aish, D., & Rogers, R. (2010). Essential leadership traits of female executives in the non-profit sector'. *The Journal of Human Resource and Adult Learning*, 6(1), 51.
- Lewis, N., Burton, J., Lewis, P., Lau, J., Stacey, C., Sydenham, E., Smith, F., Tilbury, C., Smith, T., Shlonsky, A., Parolini, Arno, & Tan, W. W. (2017). *The Family Matters Report 2017: Measuring trends to turn the tide on the over-representation of Aboriginal and Torres Strait Islander children in out-of-home care in Australia*. <http://www.familymatters.org.au/wp-content/uploads/2017/11/Family-Matters-Report-2017.pdf>
- Lunenburg, F. C. . (2011). Leadership versus Management: A Key Distinction – At Least In Theory. *Journal of Management, Business, and Administration*, 14, 5.
- Marzi, G., Balzano, M., & Marchiori, D. (2024). K-Alpha Calculator–Krippendorff's Alpha Calculator: A user-friendly tool for computing Krippendorff's Alpha inter-rater reliability coefficient. *MethodsX*, 12(102545), 102545. <https://doi.org/10.1016/j.mex.2023.102545>
- McCarthy, S., & Griffiths, L. J. (2021). How Do Leaders Enable and Support the Implementation of Evidence-based Programs and Evidence-informed Practice in Child Welfare? A Systematic Literature Review. *Human Service Organizations: Management, Leadership & Governance*, 45(5), 410–430. <https://doi.org/10.1080/23303131.2021.1929626>
- McMurray, B., Samelson, V. M. , Lee, S. H. , & Tomblin, J. B. (2010). Individual differences in online spoken word recognition: Implications for SLI. . *Cognitive Psychology*, 60(1), 1–39.
- Novins, D. K., Green, A. E., Legha, R. K., & Aarons, G. A. (2013). Dissemination and implementation of evidence-based practices for child and adolescent mental health: A systematic review. *Journal of the American Academy of Child & Adolescent Psychiatry*, 52(10), 1009–1025.

- Open Learning Initiative. (2025, May 10). *OzInspire - Evidence Based Decision Making May 2024 - May 2025*. Carnegie Mellon University. <https://oli.cmu.edu/jcourse/webui/login/required.do?redirect=%2Fjcourse%2Fwebui%2Fmyhome.do>
- Plath, D. (2012). Support for evidence-based practices in a human service organisation. *Administration in Social Work, 37*(1), 25–58.
- Podsakoff, P. M., MacKenzie, S. B., Podsakoff, N. P., & Lee, J. Y. (2003). The mismeasure of man (agement) and its implications for leadership research. *The Leadership Quarterly, 14*(6), 615–656.
- Podsakoff, P. M., & Organ, D. W. (1986). Self-Reports in Organizational Research: Problems and Prospects. *Journal of Management, 12*(4), 531–544.
- Reichenpfader, U., Carlford, S., & Nilsen, P. (2015). Leadership in evidence-based practice: a systematic review. *Leadership in Health Services, 28*(4), 298–316.
- Rocque, M., Welsh, B. C., Greenwood, P. W., & King, E. (2014). Implementing and sustaining evidence-based practice in juvenile justice: A case study of a rural state. *International Journal of Offender Therapy and Comparative Criminology, 58*(9), 1033–1057.
- Sarros, J. C., Cooper, B. K. , & Santora, J. C. (2011). Leadership vision, organizational culture, and support for innovation in not-for-profit and for-profit organizations. *Leadership & Organization Development Journal, 32*(3), 291–309.
- Shaw, K. M., Cartwright, C. , Sankaran, S. , Kelly, J., Dick, B., Davies, A., & Craig, J. (2013). Leadership in faith-based aged and community care. *Leadership in Health Services, 26*(4), 312–321.
- Shin, J., & McClomb, G. E. (1998). Top executive leadership and organizational innovation: An empirical investigation of nonprofit human service organizations (HSOs). *Administration in Social Work, 22*(3), 1–21.
- Shonkoff, J. P., Richter, L., Van Der Gaag, J., & Bhutta, Z. A. (2012). An integrated scientific framework for child survival and early childhood development. *Pediatrics, 129*(2), e460–e472.
- Simons, S. (1998). *Restoring the organization: Using team inquiry to co-construct alternative approaches to leadership within a division of a multinational, manufacturing corporation*.
- Smith, S. R., & Phillips, S. D. (2016). The changing and challenging environment of nonprofit human services: Implications for governance and program implementation. *Nonprofit Policy Forum , 7*(1), 63–76.
- Social Science Statistics. (2026, March 12). *T-Test Calculator for 2 Dependent Means*. <https://www.socscistatistics.com/tests/ttestdependent/calculator/>
- Social Work Policy Institute. (2018, October 19). *Evidence-Based Practice Social Work Policy Institute*. Socialworkpolicy.Org.

The Department of Social Services. (2025, February 23). *Protecting children is everyone's business*. Australian Government Department of Social Services.

Young, D. R. ., Hollister, R. M. ., & Hodgkinson, V. Ann. (1993). *Governing, leading, and managing nonprofit organizations : new insights from research and practice* (1st ed.). Jossey-Bass Publishers.

Yukl, G., & Lepsinger, R. (2005). *Why integrating the leading and managing roles is essential for organizational effectiveness*. 34(4), 361–375.

Zaleznik, A. (1977). Managers and leaders: Are they different. *Harvard Business Review*, 67–78.

Zhu, L., Lian, Z., & Engström, M. (2020). Use of a flipped classroom in ophthalmology courses for nursing, dental and medical students: A quasi-experimental study using a mixed-methods approach. *Nurse Education Today*, 85(104262).

11. Appendices

11-1 Ethics Approval

Professor Roslyn Cameron

Dr Fiona Chatteur

E-mail: ethics@torrens.edu.au

09/08/2024

Dear Ros and Fiona

Human Research Ethics Approval Notification

Torrens University Australia Human Research Ethics Committee (HREC)

Ethics Application 0381 Examining the Impact of Evidence-Based Decision-Making Professional Development in the Child Welfare Sector

E3.0 More than Low Risk (HREC)

Thank you for submitting 0381 for ethical review. This research project was reviewed by the HREC at the 26 June 2024 meeting.

I am pleased to advise you that 0381 meets the requirements of the National Statement on Ethical Conduct in Human Research (2023) and ethical approval

for this research project has been granted.

In this research project the methodology and participating site(s) are as stated in the application. If additional sites are engaged or withdrawn, the Principal

Investigator is required to notify the Human Research Ethics Office.

The approved documents:

Document Type File Name Date Version

CV Fiona Chatteur_academic_2024 22/05/2024 1

CV CV - Denise Rousseau 05/06/2024 1

CV CV - Dr Lisa J. Griffiths 05/06/2024 1

CV CV - Jarrod Edwards 05/06/2024 1

CV CV - Tim Lippis 05/06/2024 1

CV CV Barends 2022 05/06/2024 1

CV Cameron RA CV April 2024. doc 05/06/2024 1

Research Endorsement

FW_ Mngt Endorsement Request Ethics Application HREC#381 -

Fiona Chatteur - Outlook

06/06/2024 1

Permissions & Authorisations OzChild_JarroEdwards_Signed_20240606 06/06/2024 1

Permissions & Authorisations Cntr_EBMgt_Permission 07/06/2024 1

Participant Recruitment OzChild Email_Certificate 07/06/2024 1
Participant Recruitment OzChild Email_Certificate_2weeks 07/06/2024 1
Participant Recruitment OzChild Leaders 07/06/2024 1
Participant Recruitment OzChild Leaders_2weeks 07/06/2024 1
Participant Recruitment OZCHILD_ Participant Recruitment Flyer 07/06/2024 1
Other Project-related Fully Executed RCA_Torrens_OzChild 07/06/2024 1
Data Collection Demographics 07/06/2024 1
Signature sig 07/06/2024 1
Research Endorsement Management Endorsement
CameronChatteurGriffithsEdwards_Project 13/06/2024 1

Page 1 of 2

Document Type File Name Date Version
Data Collection Pre and post tests 04:07:24 04/07/2024 2
PICF EBDM_ PICF_CertificateCohort_4:07:24 04/07/2024 4
PICF EBDM_ PICF_Cohort2_4:07:24 04/07/2024 4
Project Description Project Description EBDM v06_04:07:24 04/07/2024 2
Data Collection Data Collection Process -Think Aloud 04/07/2024 1
Response Summary 0381_Ethics Response Summary 05/07/2024 2
Response Summary 2024-04OS Item 3.1.1 Ethics Application 0381 08/07/2024 1
Data Collection Cases Protocol Analysis 09/08/2024 1

Approval of this research project is valid from 09/08/2024 for a period of three (3) years, subject to the following conditions:

The Principal Investigator will immediately report anything that might warrant review of ethical approval of the project.

The Principal Investigator will notify the HREC of any proposed amendments to the originally approved ethics application. An Amendment Request is

to be submitted, and any change must be approved prior to implementation.

The Principal Investigator will notify the HREC of any plan to extend the duration of the project past the approval period. An Extension Request is to be

submitted, and any change must be approved prior to implementation.

The Principal Investigator will notify the HREC of any complaint received regarding ethical aspects of the project, in a timely manner.

The Principal Investigator will notify the HREC of any adverse or serious events that may impact participant welfare, within 48 hours of occurrence.

The Principal Investigator will provide an Annual Report to the HREC and provide a Final Report to notify the HREC when the project is completed.

The Principal Investigator will notify the HREC if the project is discontinued at a participating site before the expected completion date.

The Principal Investigator will notify the HREC of their inability to continue as Principal Investigator including the name of and contact information for a replacement.

The Principal Investigator will comply with audit requests or other forms of monitoring by the HREC, as required.

This letter constitutes ethical approval only. This research project cannot proceed at any site until site specific authorisation or relevant permission, has been

obtained from the appropriate authority.

Should you have any queries about the review of your research project please contact the Human Research Ethics Office, ethics@torrens.edu.au.

The Torrens University Australia HREC wishes you every success in your research.

Kind regards

Torrens University Australia HREC

The Torrens University Australia HREC is constituted and operates in accordance with the *National Statement on Ethical Conduct in Human Research (2023)*

11.2 Pre- and Post-tests

3/1/26, 4:14 PM

EBM Course Pretest

EBM Course Pretest

Schedule: [Update](#)
Available now.

Page 1

Question 1

Read the following case and determine which source(s) of evidence was consulted. Check all that apply.

Select all that apply.

10 points

To lower costs and increase revenues, the boards of two academic medical centers have decided to merge. Together they account for 16 percent of the market for medical care. According to a well-known consulting firm, each percentage increase in market share would bring the new organization an additional \$50 million of revenue. Both CEOs agree that the new organization would reduce administrative overhead by 20 percent. They also argue that it is common knowledge that large medical centers tend to have better patient outcomes than do small medical centers. A survey of the patients, commissioned by the boards, however, indicates for various reasons, a large majority is opposed to the merger.

- A. Evidence from the scientific literature
- B. Evidence from the organization
- C. Evidence from stakeholders
- D. Evidence from practitioners
- E. I don't know

Question 2

You have recently been appointed as managing director for a small manufacturing firm. In your first months working there, it becomes clear to you that the accounting system is not working properly and there is limited insight into the firm's fixed and variable costs. The financial director suggests introducing activity-based costing, a widely used costing method. Since you are not sure whether activity-based costing will also work in small firms, you decide to consult evidence from a number of sources before you decide to implement this method.

Click and drag each box to order them. Tab: Select next item. Shift+Tab: Select previous item. Arrow Down: Move item down. Arrow Up: Move item up.
10 points

Please read the following four evidence sources and rank the evidence according to its strength/trustworthiness (most trustworthy evidence on top).

The financial director shows you an article in a popular management magazine describing the results of a study involving the implementation of activity-based costing in three small manufacturing companies. The result of this study is based on interviews with four financial directors. Three of them are positive about the benefits of activity-based costing. He therefore advises you to implement the method.

You contact a senior consultant at a well-known consulting firm. This consultant tells you that, based on his 15 years of experience with a large

number of small manufacturing firms, he judges activity-based costing to be too expensive and complex for small and medium-sized businesses. He therefore advises you not to implement the method.

Findings from a survey regarding the use of activity-based costing, published in a popular magazine, indicate that 53 percent of all multinational companies in the United States have implemented this method.

A recent study published in an academic journal shows that activity-based costing has benefits for small and medium-sized enterprises. The study compares 15 small organizations that have implemented the model with 19 small organizations that have not implemented it. Objective performance measures before and after implementation suggest that activity-based costing leads to a better understanding and control of costs.

Question 3



Select one answer.

10 points

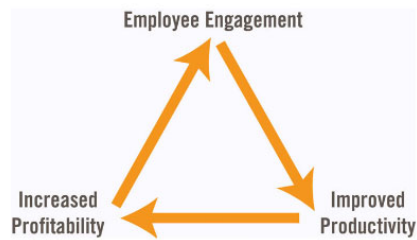
Read the following scenario and determine whether the problem is sufficiently clear:

The passage of the Affordable Care Act (ACA, also called Obamacare) in 2010 required that US hospitals report their performance on new metrics. These new metrics allow the federal government to reimburse hospitals for the costs of care based on their performance quality.

Riverview Healthcare, a large hospital located in the eastern United States, decided to take the challenge of implementing the new performance metrics one step further by also changing the financial compensation plan for its physicians to improve the quality of care. The new plan would compensate physicians with both a base salary and a variable bonus tied to their individual performance on the ACA metrics. According to Riverview's executives, the new plan would give greater emphasis to quality over quantity of care and result in a more patient-centered view.

How clearly defined is the problem (what, who, when, where, why)?

- A. Very clear
- B. Fairly clear
- C. Somewhat unclear
- D. Very unclear
- E. I don't know



Select one answer.

10 points

Read the following scenario and determine the PICOC terms:

The organizational data of a German pharmaceutical company indicate that the firm's overall productivity is below the average in the sector. The HR director suggests conducting an employee engagement survey, arguing that a company's productivity is dependent on the collective performance of its employees. After all, according to Gallup—a global analytics and consulting firm—engagement is the most important driver of performance by employees. The outcome of the survey shows that the company's engagement score is below Gallup's Global Employee Engagement Benchmark. For this reason, structured interviews with employees are conducted to identify factors that negatively affect their engagement. Based on the outcome of the interviews, several techniques for improving employee engagement are presented to the board, and a corporate action plan is agreed upon.

What is the outcome of interest?

- A. The company's overall productivity
- B. Employee performance
- C. Employee engagement
- D. The collective performance of the employees
- E. I don't know

Question 5**Read the following scenario:**

Select one answer.

10 points

A company decides to survey its clients to understand their satisfaction with the new service the company has been providing for two months. For this purpose, the company's 40 largest clients are invited to complete an online survey about their experiences.

Which of the following might negatively affect the outcome of the survey?

- A. A sample of 40 is too small.
- B. An online survey is not a good method to measure customer experience.
- C. The sample is not random.
- D. The period of two months after the introduction of the new service is too short.
- E. I don't know

Question 6**Are the following four questions and statements adequately formulated? Select Yes or No.**

Fill in the blanks by selecting one option from each menu.

How likely are you to go out for dinner and a movie this weekend?

Part 1: 2 points

Part 2: 2 points

-- ▼

In the team meetings I attend, risk management is always on the agenda.

-- ▼

The last team meeting was well attended.

-- ▼

How would you rate the new procedure that was implemented last month that has dramatically reduced the number of errors?

-- ▼

Part 3: 2 points

Part 4: 2 points

8 points

Question 7

Over a five-year period, whose professional expertise would you judge to be the least trustworthy (valid and reliable)?

Select all that apply.

10 points

- A. A car salesperson specializing in selling secondhand cars
- B. A management consultant specializing in culture change
- C. An eye surgeon specializing in eye laser surgery
- D. A baker specializing in making sourdough bread

Question 8

Read the following scenario.

Select one answer.

10 points

Your CEO is very enthusiastic about a new, revolutionary model developed by Google to stimulate employees' creativity and considers implementing it in the organization (an international bank). She suggests paying a visit to Google's headquarters in California to see if this new model is indeed as successful as claimed. However, if your CEO were to visit Google, she might hear only success stories that will confirm her prior (positive) beliefs regarding the model.

What would be the best advice you could give your CEO to prevent confirmation bias?

- A. Cancel the visit and instead invite people from Google to give a presentation.
- B. Write down in advance many critical questions that should be asked during the visit.
- C. In addition to visiting Google, visit a company where the implementation of the model has not led to greater creativity.
- D. Ask Google for hard, objective evidence that the model is a success.
- E. Bring along someone who can play devil's advocate and take an opposite view.

Question 9

Read the following scenario.

Select all that apply.

10 points

At a board meeting of a large British software development firm, the low number of new products and innovations is discussed—after multiple restructurings, the company performs far below the average for its sector. The HR director of the firm explains to his colleagues that companies with a more ethnically diverse workforce tend to perform better. He points out that the company's software design teams tend to be dominated by white, 25-year-old men with degrees from MIT. The CEO, however, thinks that diverse teams tend to have more communication problems and task conflicts and therefore prefers to maintain the status quo.

In this scenario, what are the TWO most important PICOC terms to use in a search for empirical studies?

- A. P:
Software design teams
- B. I:
Ethnic diversity
- C. C:
Status quo
- D. O:
New products/innovation
- E. C:
A British software company that has experienced multiple restructurings
- F. I don't know

Question 10

Have a look at the screenshot below. A researcher has combined multiple searches with the term *ethnic diversity* and two related terms. This person combined the outcome with OR and then applied the filter to identify the meta-analyses and/or systematic reviews.

Select one answer.

10 points

<input type="checkbox"/> Select / deselect all <input type="button" value="Search with AND"/> <input type="button" value="Search with OR"/> <input type="button" value="Delete Searches"/>				
	Search ID#	Search Terms	Search Options	Actions
<input type="checkbox"/>	S6	S4 AND S5	Search modes - Boolean/Phrase	View Results (4)
<input type="checkbox"/>	S5	TI(meta-analy*) OR AB(meta-analy*) OR TI("systematic review") OR AB("systematic review")	Search modes - Boolean/Phrase	View Results (7,505)
<input type="checkbox"/>	S4	S1 OR S2 OR S3	Search modes - Boolean/Phrase	View Results (291)
<input type="checkbox"/>	S3	TI "team diversity" AND AB "team diversity"	Limiters - Scholarly (Peer Reviewed) Journals Search modes - Boolean/Phrase	View Results (28)
<input type="checkbox"/>	S2	TI "cultural diversity" AND AB "cultural diversity"	Limiters - Scholarly (Peer Reviewed) Journals Search modes - Boolean/Phrase	View Results (188)
<input type="checkbox"/>	S1	TI "ethnic diversity" AND AB "ethnic diversity"	Limiters - Scholarly (Peer Reviewed) Journals Search modes - Boolean/Phrase	View Results (95)

Click image to enlarge.

As you can see, the final result is 4. However, it is likely that some relevant meta-analyses were missed. Why?

- A. The person should have searched in a different database.
- B. The person forgot to select the Peer Reviewed option.
- C. In S6, the person should have used OR instead of AND.
- D. In S1, S2, and S3, the person should have used OR instead of AND.
- E. I don't know

Question 11

Read the following abstract of an empirical study, "Leadership Styles of Entrepreneurs and Managers."

Select one answer.

5 points

This study tested hypotheses related to leadership styles of entrepreneurs and managers. A random sample of 621 owners of small businesses and 127 managers of large companies participated in a survey conducted in early 2014. A questionnaire was used for data collection and was sent out to all participants for them to rate their leadership style. We found significant differences in leadership styles between the two groups. Entrepreneurs were more likely than managers to use inspirational and charismatic leadership behaviors. Managers had higher levels of passive, laissez-faire leadership behaviors than did entrepreneurs.

What is the research design of this study?

- A. A cross-sectional study
- B. A randomized controlled study
- C. A before-after study
- D. A meta-analysis or systematic review
- E. I don't know

Question 12

Read the following scenario.

Select one answer.

10 points

Search Inside Yourself (SIY) is a mindfulness program that was developed by Google. It consists of an interactive two-day program followed by four weeks of daily 30-minute online sessions. In the past years, more than 200 companies in 20 countries have offered the program to their employees. The program claims that it reduces stress, improves focus, raises performance, and improves interpersonal relationships.

A researcher decides to see whether these claims stand up to scientific scrutiny. She selects a random sample of 60 managers from the UK who were enrolled in the program. The results of the study demonstrate that the SIY program is effective: after four weeks, the participants showed statistically significant lower stress levels, a better focus, and a higher task performance ($p < .05$).



Could the outcome of the study be affected due to the placebo effect? Why or why not?

- A. Yes, because there is no comparison with a random sample of employees who continued their daily work.
- B. Yes, because there is no comparison with a random sample of employees who did something relaxing for 30 minutes (e.g., listened to music, took a stroll).
- C. No, because there is a pre- and post-measurement, so it is not likely that a placebo effect has affected the outcome.
- D. No, because the results are statistically significant at the $p < .05$ level, which indicates the effect is likely to be genuine.
- E. Yes, because it is unclear whether the effect sizes are practically relevant.
- F. I don't know

Question 13



Select one answer.

10 points

Read the following abstract from a study, “Preliminary Investigation of Employee’s Dog Presence on Stress and Organizational Perceptions”:

A pre-post design was used to compare differences between employees who bring their dogs to work (DOG group) and employees who do not bring their dogs to work (NODOG group) on perceived stress. Significant differences were found between the two groups. Although perceived stress was similar at baseline, over the course of the day, stress declined for the DOG group with their dogs present and increased for the NODOG group. The NODOG group had higher stress than the DOG group by the end of the day ($p < 0.04$, $d = .03$). In addition, a significant difference was found in the stress patterns for the DOG group on days their dogs were present and absent. On dog absent days, owners’ stress increased throughout the day, mirroring the pattern of the NODOG group ($p < 0.02$, $d = .04$).

Based on these findings, what would your conclusion be?

- A. The effect of dogs at the workplace on perceived stress is very small and practically irrelevant.
- B. Not allowing employees to bring their dog to work induces stress.
- C. The statistical significance of the findings is below .05, which is too low.
- D. Managers should consider allowing employees to bring their dogs to work, as this substantially reduces stress.
- E. I don't know

Question 14

Read the following scenario.

Select one answer.

15 points

A recent study published in the Journal of Accounting, Finance & Management Strategy found that women with an interest in fashion are more likely than women without a fashion interest to overuse their credit cards.

Download a copy of the study << here >>

Given the study’s design and methodological quality, how would you qualify the trustworthiness of this finding?

[Click here to use the grading table.](#)

- A. High (90%)
- B. Moderate (80%)
- C. Limited (70%)
- D. Low (60%)
- E. Very low (< 60%)
- F. I don't know

Question 15

Consider the following scenario:

Select one answer.

15 points

During a board meeting of a large German automotive manufacturer the report of an external consultant is discussed. The report states that in the last quarter the number of sales has decreased relative to its competitors. The CEO warns that this could seriously damage the company's market share, and therefore demands immediate action. The company's head of marketing thinks the decrease is due to the unfavorable economic situation, so she suggests lowering the prices of the luxury cars. The head of sales, however, thinks that the decrease is caused by the high number of young, inexperienced sales agents, so he suggests sending them to a training course to improve their sales techniques. Finally, the Chief Operation Officer thinks that the decrease can be attributed to a decline in production that is caused by the implementation of a new production process.

What evidence from the organization should the board collect first?

- A. External data regarding customers' perception of prices of luxury cars and whether its sales has decreased in the past quarter.
- B. Internal data regarding the number of young, inexperienced sales agents and whether this number has increased in the past quarter.
- C. Internal data regarding the number of cars produced and whether there is a decline that coincides with the implementation of the new production process.
- D. External data regarding the average number of sales in the sector and whether the company's sales have decreased relatively in the past quarter.
- E. Internal data regarding the number of sales agents that have received a sales training.
- F. I don't know

Question 16

Read the following scenario.

Select all that apply.

5 points

Compared to the previous month, hospital X has treated 100 more patients, whereas hospital Y has treated 1,000 more patients.

Check the 3 types of contextual information you would need to determine which hospital is doing better from a financial perspective.

- A. The total number of patients each hospital treats per month
- B. The medical departments/specialties that gained the most patients
- C. Each hospital's profitability per month
- D. Each hospital's average patient satisfaction score
- E. The profitability of each hospital's "average" patient

Question 17

Read the following scenario.

Select one answer.

10 points

You are responsible for the organizational data of the National CPR Institute, which offers CPR (cardiopulmonary resuscitation, also known as cardiac massage) training courses through 1,200 local training centers. Over the past years the average pass rate for this training is 85 percent. However, the Institute's information system shows that in the past month at 67 training centers, the pass rate was below 80 percent. The institute's medical director asks you how she should interpret these numbers. When you look at the characteristics of these 67 centers, you notice that most of them are small centers that provide training to relatively small groups.

Based on this information, what would be your answer to the medical director?

- A. The pass rate at the 67 training centers is below average, but the cause is unclear.
- B. The pass rate at the 67 training centers is not too low; it is probably average.
- C. It is unclear whether the pass rate at the 67 training centers is below average.
- D. I don't know

Question 18**In the scenario above (Question 17), what would you recommend the medical director?**

Select one answer.

10 points

- A. Wait three months to see whether the pass rates of the 67 centers have changed.
- B. Pool data from several small centers to develop a larger data set.
- C. Calculate and report a 95% confidence interval.
- D. All of the above.
- E. I don't know

Question 19**Read the following scenario.**

Select one answer.

10 points

An international food retailer considers launching a new type of business: high-end grocery stores stocking the best and freshest organic health foods and a wide selection of prepared gourmet food. The company's marketing director suggests that these stores should be located in high-income areas, because empirical research indicates there is a correlation between income and food expenditure. The company's data analyst decides to analyze the organizational data by regressing level of income (predictor metric) on monthly food expenditure (expressed in units of \$10). The analyst finds a regression coefficient of $b = 8.7$.

What does this mean?

- A. This means that the income does not predict food expenditure.
- B. This means that for every (one) greater unit of income, on average \$87 more per month is spent on food.
- C. This means that for every (one) greater unit of income, on average \$8.7 per month more is spent on food.
- D. This means that it remains unclear whether income predicts food expenditure.
- E. I don't know

Question 20

Which TWO factors determine the trustworthiness of evidence acquired from stakeholders?

Select all that apply.

10 points

- A. The extent to which the evidence could be affected by cognitive biases.
- B. The extent to which the evidence is representative for all stakeholders
- C. The extent to which the stakeholders could freely express their views and feelings
- D. The extent to which the evidence is based on objectifiable facts
- E. I don't know

Question 21

Read the following scenario:

Select one answer.

10 points

A large multinational company decides to survey its most important stakeholders – the customers – to understand their satisfaction with the new service the company has been providing for two months. For this purpose, the company's 40 largest clients are invited to complete an online survey about their experiences.

Which of the following might negatively affect the trustworthiness of the outcome?

- A. A sample of 40 is too small.
- B. An online survey is not a good method to collect stakeholder evidence
- C. The period of two months after the introduction of the new service is too short.
- D. The 40 largest clients may not be representative for the company's total population of clients.
- E. I don't know

Question 22

Read the following scenario.

Select one answer.

10 points

A young entrepreneur presents his business plan for an IT start-up to a group of early-stage investors: people who provide funding for start-up businesses still in the development stage. The investors, mostly former investment bankers or venture capitalists with relevant experience in a wide range of industries, get very excited about the business plan. Based on their experience and professional judgment, they estimate that the probability that this start-up will be successful is higher than 50 percent.

Do you agree with this estimate?

- A. Yes, because investors should be able to identify a good business plan (it's their job).
- B. Yes. The investors have relevant experience in a wide range of industries, so they should be able to determine whether a start-up is likely to be successful.
- C. No, because the prior probability that an IT start-up will survive the first five years is less than 10 percent - you can check this for yourself on Google
- D. No. There is no evidence available that supports or contradicts the investors' estimate, so the probability should be estimated at 50/50.

Question 23

Read the scenario below.

Select one answer.

10 points

A tech start-up in Silicon Valley (California, USA) wants to improve the performance of its engineers and software and R&D teams. The HR director suggests increasing employee engagement because the findings of a new Harvard Business Review Analytic Services report of more than 550 tech company executives indicate that three-quarters of those surveyed said that most employees in their organizations are not highly engaged. When you consult the research literature you find several studies that suggest the correlation between employee engagement and performance is between 0.12 and 0.26. In addition, you find a recent systematic review of studies on drivers of knowledge work performance. The study included a table (shown below) listing factors and their correlation with performance.

Would you recommend increasing the company's employee engagement?

Factor	Performance measure	Number of studies	Mean correlation
Intra-team trust	Team performance	200+	0.62
Social cohesion	Team performance	40+	0.59
Supervisory support	Employee performance	50+	0.53
Information sharing	Team performance	60+	0.51
Goal clarity	Team and employee performance	80+	0.49

- A. Yes
 B. No
 C. I need more information.

Question 24



Select one answer.

10 points

Read the following scenario.

A large German hospital chain with 42 locations decides to introduce a new compensation plan for its physicians in order to improve patient care. In this new plan, physician compensation would include both a base salary and a bonus portion tied to two metrics: low re-admissions of the same

patient for the same condition and the average satisfaction score of the physician's patients. To assess the effect of the new plan, the hospital's chief medical officer (CMO) decides to conduct an after-action review (AAR) with a representative group of 20 physicians. During the review, the CMO focuses on learning and improvement rather than on evaluation or judgment. This developmental, nonpunitive focus yields a lot of honest and straightforward feedback: most physicians are skeptical about the need for the new compensation plan and whether the change in incentives has improved the hospital's care quality. In addition, they question whether the hospital's management has evidence to support its claim that the introduction of the new compensation plan has improved patient care.

What is the BEST advice you could give to the CMO about how to improve the reliability of the assessment of the effect of the new compensation plan?

- A. Increase the number of participants in the after-action review and include people who can bring in other perspectives (e.g., nurses and/or patients).
- B. Conduct a pilot in one location: Introduce the new compensation plan and closely monitor whether the number of readmissions declines and patient satisfaction improves.
- C. Conduct a randomized controlled trial: Randomly assign the 42 locations to an intervention group (new compensation plan) and a control group (old compensation plan) and measure the number of readmissions and patient satisfaction a few weeks before and several months after the trial.
- D. Increase the number of participants in the after-action review and include input from at least one additional source of evidence (e.g., organizational data).
- E. I don't know

11.3 "Think aloud" scenarios

Pre Intervention Scenario:



EBM CASE AZIZA

Rev. August 2024

AZIZA

Aziza Loyalty Solutions Inc. provides customer loyalty programs designed to retain customers by rewarding their repeat business. Examples include Starbucks Rewards, Amazon Prime, and Hilton Honors. The company operates with about 2,500 employees across 15 offices worldwide and has an annual revenue of \$2 billion USD.

The HR director feels that the company's workforce is aging, mainly consisting of employees in the age range of 30 to 65. He points out that younger employees do not tend to stay long at Aziza, which he attributes to the inability of team leaders to manage generational differences: *"Younger generations such as Millennials and Gen Zers relate better to a coaching style of management than to a more traditional top-down authoritative approach. In addition, they prefer a more collaborative and in-person way of interaction."* He therefore proposes offering all team leaders a workshop on generational differences in the workplace.

The company's CFO, however, claims that generational differences are a myth and should not be taken too seriously: *"All people are different, and a good team leader with strong people management skills should be able to take that into account."* Finally, a senior executive points out that nowadays, young employees tend to explore multiple career paths and, on average, stay less time with one employer: *"This is just one of those facts of life that we have to deal with."*

How would you take an evidence-based approach to the (assumed) problem and its preferred solution(s)?

Post-Intervention Scenario:



EBM CASE PSY-Care

Rev. July 2024

PSY-Care

Meeting the Demand: Overcoming Production Shortfalls



PSY-Care is a large mental health care organization with 27 locations. The organization has more than 20,000 employees and offers psycho-medical treatment for people with anxiety disorders, depression, eating disorders, personality problems, and PTSD. After rapid growth in the first few years, the organization's productivity now seems to be below the targets agreed upon with the health insurance companies that fund the organization. If these targets are not met by the end of the year, the insurance companies will cut the organization's funding, leading to job losses.

The board of directors believes that the productivity targets are not being met because therapists are not seeing enough patients. They have drafted a letter intended for all the organization's healthcare professionals, urging them to cancel all non-patient-related activities (e.g., training activities, team events, policy meetings) and see more patients. The professionals, however, point out that their workload is already very high and that something else must be going on. Some managers suggest that it is most likely an administrative problem, like too much paperwork or poor matching of therapists to patients.

How would you take an evidence-based approach to the (assumed) problem and its preferred solution(s)?

11.4 Keyword Lists

Asking		
Asking (determining problem)	Determine problem	Identify assumptions
Evidence for problem	Is the problem clear?	claim
Confirm problem		

Acquiring		
Acquiring or getting Evidence	What is the Evidence for solution	Does evidence support solution?
Look at the figures	Capturing data	gathering

Search (sub-code for Acquiring)		
Research database	Extended search	Google scholar
Thesaurus	Search terms	research
Narrowing search results	Academic terms	
Methodological filters	Testing search terms	

Organisational Data (sub-code for Acquiring)		
Organizational decisions	Organisational evidence/data	resources (in organisation)
Internal data	Exit surveys	Employee surveys
demographics	statistics	Organisational context

Appraising		
Peer-reviewed	Quality of evidence	Appraising evidence
Sound evidence	trustworthiness	Providing evidence of decision
Consequences (to organisation)	cause	
trustworthy	expertise	Relevant/relevance
effect	Appropriate/appropriateness	Correlate/correlation
precise	measure	Confidence/confidence interval
Measure/measurement	evaluate	bias
Dig deeper	Judgement	unfair
Representative/ness	Cognitive bias	Group characteristics
reliability	Citation (of research)	Base statistics

Scientific Evidence (sub-code for appraising)		
Journal	Longitudinal studies	Empirical Studies
Scientific evidence	Controlled studies	baseline
Random sample	generalisable	Randomised controlled trial
Control group	Statistics (external to org)	demographics
Experimental group		

Stakeholders

Stakeholders	values	Stakeholder agreement with problem
feelings	concerns	Secondary stakeholders
influence	Primary stakeholders	Relevant stakeholders
affected	interests	beliefs
Listening to stakeholders	affected	Informed consent
ethics	Listening to stakeholders	power
Stakeholder evidence	views	Stakeholder evidence
societal	political	subjective
colleagues	perceptions	bosses
clients	Employees (depending on context – not practitioners)	leader
	Qualitative data (from organisation) e.g. surveys	

Practitioners		
experience	practitioners	Community of practice
judgement	Practitioner agreement with problem	therapists
professional	Subject matter expert	Team (team members)
Employees (who are not bosses)		

Applying		
Applying evidence	context	justice
solution	Expected value	costs
benefits	probability	risk
beneficence	respect	implementation
Practical relevance		
group		

Culture		
Principles	leadership	Critical thinking
support	Organisational culture	Management style

Self reflection by managers	Personality types	strengths of the organization
supportive	norms	Organisational culture
culture	skills	training

Aggregating (gathering data together)		
Aggregating	Multiple sources of evidence	Documenting/documentation
Industry comparisons	Industry benchmarking	odds
Weighing evidence	Best available evidence	likely
Probability/probable	true	Likelihood (of evidence)
chance	Bayes rule	Comparing with Other companies
Prior (probability)	sector	

Assessing (the outcome)		
Assessing outcome	Decision justification	Outcomes
Executed/execution	intervention	impact
implementation	Assess/assessing	consequences
Routine/nonroutine decision	effectiveness	procedures
Novel decisions	rules	Implementation fidelity
training	dose	Cause and effect

The 6As. Six steps characterize the processes applied in EBM: Asking (what problem or solution to address), Acquiring relevant evidence, Appraising its quality, Aggregating across evidence sources, Applying the evidence to make a decision, and Assessing the outcome.

4 Pillars: Organisational data, practitioners, stakeholders' voices and concerns, scientific literature (studies)

11.5 Demographic Survey

Q1 This research project involves "think aloud" protocol analysis exercises. In order to do this we need to contact you again in the future. Please provide a name and email address we can contact you on. This information is confidential and will not be shared with third parties.

Q2 Name:

Q3 Email:

Q4 What is your gender?

Male

Female

Non-binary/ third gender

Prefer not to say

Prefer to self describe:

Q4 What is your age?

20-29 years

30-39 years

40-49 years

50-59 years

60 years and over

Q5 How many years of professional work experience do you have?

0-2 years

3 to 5 years

6 to 10 years

11 to 20 years

Over 20 years

Q6 What post-secondary qualifications do you have (Choose all that apply)

Doctoral degree

Master's degree

Graduate diploma

Graduate certificate

Bachelor's degree

Associate degree

Diploma

Certificate IV

Certificate 1-3

Other:

Q7 Do you have previous experience in any of the below (Choose all that apply)

Human Resources

Human Resource Development

Leadership of business units

Retail

Sales

Marketing

Product Innovation

Psychology

Psych Nursing

Allied Health

Other:

We thank you for your time spent taking this survey.

Your response has been recorded.

11.6 Krippendorff Alpha Calculations (Technical notes)

Participant 1 Pre-"think aloud"

Participant 1 (P1) was in the Control group.

Number of utterances =45

Number of codes = 12

Number of coders = 2

Columns 1-12 = Coder 1

Columns 13-24 = Coder 2

Values 1 = Category assigned 0 = not assigned

Using ChatGPT (Open AI, 2026) to do the analysis the data returned a reliability of **0.72** for Krippendorff's alpha.

This was broken down as:

Per-Code Krippendorff's α (Nominal)

Code	Alpha
1 Asking	0.74
2 Acquiring	0.68
3 Search	0.81
4 Organisational data	0.59
5 Appraising	0.77
6 Scientific Evidence	0.72
7 Aggregating	0.65
8 Stakeholders	0.83
9 Applying	0.70
10 Culture	0.76
11 Assessing	0.79
12 Practitioners	0.61

An alpha value of **0.72** is considered good for Krippendorff's alpha, which affords tentative conclusions, but is not considered highly accurate. With multi-label nominal data it is difficult to achieve a high level of accuracy. For this type of data, it is considered a solid reliability score.

Participant 12 Pre-"think aloud"

Number of utterances =25

Number of codes = 12

Number of coders = 2

Columns 1-12 = Coder 1

Columns 13-24 = Coder 2

Values 1 = Code assigned 0 = not assigned

Using ChatGPT (Open AI, 2026) to do the analysis the data returned a reliability of **0.58** for Krippendorff's alpha. This is broken down into the codes (categories):

Per-Code Krippendorff's α (Nominal)

Code	Alpha
1 Asking	0.34

Code	Alpha
2 Acquiring	0.50
3 Search	1.00
4 Organisational Data	0.42
5 Appraising	0.06
6 Scientific Data	1.00
7 Aggregating	0.33
8 Stakeholders	0.86
9 Applying	-0.02
10 Culture	0.46
11 Assessing	1.00
12 Practitioners	1.00

Overall Reliability

Mean α across 12 categories

$$\alpha_{mean} = 0.58$$

One reason for the low reliability score is the low number of utterances for this participant. This is a low alpha value, and a weaker agreement is indicated. Because the number of utterances is so small, it is difficult to get a higher reliability score, which also indicates low agreement between the coders.

P7 Post "think aloud"

Number of utterances =29

Number of codes = 12

Number of coders = 2

Columns 1-12 = Coder 1

Columns 13-24 = Coder 2

Values 1 = Category assigned 0 = not assigned

Using ChatGPT (Open AI, 2026) to do the analysis the data returned a reliability of **0.71** for Krippendorff's alpha. This is broken down into the categories:

Per-Code Krippendorff's α (Nominal)

Codes	Alpha
1 Asking	0.55
2 Acquiring	0.48

Codes	Alpha
3 Search	1.00
4 Organisational data	0.60
5 Appraising	0.73
6 Scientific data	1.00
7 Aggregating	0.69
8 Stakeholders	0.71
9 Applying	0.52
10 Culture	0.64
11 Assessing	0.82
12 Practitioners	0.78

Overall Reliability

Mean α across 12 categories

$$\alpha_{mean} = 0.71$$

An alpha value of 71 suggests an acceptable level of agreement, which is indicative, but not highly accurate. Tentative conclusions can be drawn from this data. For this type of data, which is multi-label nominal data, it is considered a solid reliability score.

P8 Post "think aloud"

Number of utterances =30

Number of codes = 12

Number of coders = 2

Columns 1-12 = Coder 1

Columns 13-24 = Coder 2

Values 1 = Category assigned 0 = not assigned

Using ChatGPT (Open AI, 2026) to do the analysis the data returned a reliability of **0.79** for Krippendorff's alpha. This is broken down into the categories:

Per-Code Krippendorff's α (Nominal)

Code	Alpha
1Asking	0.87
2 Acquiring	0.81

3	Search	1.00
4	Organisational Data	0.46
5	Appraising	0.55
6	Scientific Data	-0.05
7	Aggregating	-0.03
8	Stakeholders	0.91
9	Applying	1.00
1 0	Culture	1.00
1 1	Assessing	1.00
12	Practitioners	1.00

Result:

$\alpha = 0.791$

Acceptable to near-reliable agreement

Krippendorff's alpha for participant 8 was **0.79**, which is considered good coder reliability, given the multi-label nominal coding.

Given that the average of all of the Krippendorff alpha for the sample "think aloud" data returned an average of 0.7, which is considered acceptable for multi-label data, it was decided to analyze the data based on these results. The caveat is that those with a smaller number of utterances will be considered less accurate.

11.7 Pre-Post Test Results by Learning Outcome

N = 899 students. Sorted by gain (descending). Green = gain \geq 30%; Red = gain < 10%.

Question	Learning outcome (skill)	N	Pretest (%)	In course test (%)	Gain (%)	Cohen's d	95% CI
Q 10	Determining which sources of evidence were consulted in a decision-making process	843	38.1	76.3	38.2	0.62	[0.54, 0.69]
Q 20	Identifying factors that may affect the trustworthiness of evidence from stakeholders	748	19.9	57.1	37.2	0.61	[0.53, 0.69]
Q 12	Determine whether bias, confounding or effect modification may have affected the results	834	39.3	76	36.7	0.6	[0.53, 0.68]
Q 09	Determining the most important PICOC terms	420	51.7	87.9	36.2	0.77	[0.66, 0.88]
Q 13	Assessing whether a statistically significant finding is of practical relevance	813	26.4	59.8	33.3	0.54	[0.47, 0.62]
Q 17	Determine whether there could be a small number problem	746	24.7	57.5	32.8	0.53	[0.45, 0.60]
Q 23	Determining whether a decision/intervention gives you the biggest bang for your buck	751	25.6	57.4	31.8	0.52	[0.44, 0.59]
Q 14	Assessing a study's methodological quality on the basis of pre-determined criteria	813	15.6	43.7	28.2	0.59	[0.52, 0.67]
Q 19	Interpreting a correlation or regression coefficient	746	31.4	59.5	28.2	0.46	[0.38, 0.53]
Q 15	Determining what organizational evidence to acquire	766	45.8	72.6	26.8	0.44	[0.36, 0.51]
Q 24	Suggesting ways to improve the validity and reliability of an outcome assessment	717	62.6	88.6	25.9	0.45	[0.38, 0.53]
Q 07	Assessing whether professional experience is valid and reliable	854	34.9	59.7	24.8	0.4	[0.33, 0.47]
Q 11	Identifying a study's research design	834	61.5	84.4	22.9	0.4	[0.33, 0.47]
Q 08	Identifying ways cognitive biases can be overcome	854	63.8	85.8	22	0.4	[0.33, 0.46]
Q 04	Applying the PICOC method to make explicit the organizational context	886	45.1	65.6	20.4	0.34	[0.27, 0.41]

Question	Learning outcome (skill)	N	Pretest (%)	In course test (%)	Gain (%)	Cohen's d	95% CI
Q 03	Determining whether the (assumed) problem is sufficiently clear	886	33.7	52.8	19.1	0.37	[0.30, 0.44]
Q 22	Assessing the impact of a prior probability	735	26.9	45.9	18.9	0.32	[0.25, 0.39]
Q 18	Identifying a small number problem	746	68.6	86.2	17.6	0.43	[0.35, 0.50]
Q 05	Determining how to prevent selection bias when acquiring evidence from practitioners	848	73.9	89.7	15.8	0.33	[0.27, 0.40]
Q 06	Formulating clear, unambiguous, and unbiased questions	848	62.6	74.9	12.3	0.35	[0.29, 0.42]
Q 21	Determining whether selection bias may have affected the acquisition of stakeholder evidence	748	80.7	90.4	9.6	0.21	[0.13, 0.28]
Q 16	Assessing whether organizational data are relevant	746	8.2	16.2	7.9	0.18	[0.11, 0.25]
Q 02	Assessing (coarsely and in general terms) the quality of evidence	878	31.7	33.9	2.3	0.04	[-0.03, 0.10]
Q 01	Determining which sources of evidence were consulted in a decision-making process	878	10.7	11.6	0.9	0.02	[-0.04, 0.09]

Note: Cohen's d based on paired differences.